



Jeremias José Jasse

**Uma Meta-Avaliação de Questões de Escolha
Múltipla em Estatística**



Jeremias José Jasse

Uma Meta-Avaliação de Questões de Escolha Múltipla em Estatística

Dissertação apresentada à Universidade de Aveiro para cumprimento dos requisitos necessários à obtenção do grau de Mestre em Matemática e Aplicações, realizada sob a orientação científica de Adelaide de Fátima Baptista Valente Freitas e coorientação científica de João Pedro Antunes Ferreira da Cruz, Professores Auxiliares do Departamento de Matemática da Universidade de Aveiro.

Aos meus pais (*in memoriam*)

À Culentiny

Aos meus filhos

Sem eles, nada seria possível.

o júri / the jury

presidente / president

Doutor Pedro Filipe Pessoa Macedo

Professor Auxiliar do Departamento de Matemática da Universidade de Aveiro

Doutor Milton Severo Barros da Silva

Professor Auxiliar Convidado do Departamento de Epidemiologia Clínica, Medicina Preditiva e Saúde Pública da Faculdade de Medicina da Universidade do Porto

Doutora Adelaide de Fátima Baptista Valente Freitas

Professora Auxiliar do Departamento de Matemática da Universidade de Aveiro (orientadora)

agradecimentos

À Professora Doutora Adelaide Valente Freitas, minha orientadora e ao Professor João Pedro Cruz, meu coorientador pelo apoio incondicional na orientação desta dissertação, empenho, disponibilidade, discussões e sugestões à altura do tema. A sua ajuda tornou possível a realização deste trabalho.

Ao Governo de Moçambique que através do Instituto de Bolsas de Estudos disponibilizou o apoio financeiro.

Aos professores do Departamento de Matemática, em particular os do Mestrado em Matemática e Aplicações, pela sábia transmissão de conhecimentos e esclarecimentos oportunos, em todas as unidades curriculares do curso.

À Flora Samuel Jasse, minha amada esposa, pelo amor, carinho, dedicação, paciência imensurável e que soube, não obstante a distância que nos separava, estar sempre presente dando conforto e apoio necessários para superar todos os desafios e retornar aos seus braços.

Aos meus irmãos e minha sogra por estarem presentes em minha vida nos momentos em que mais os precisei.

Aos colegas do curso, compatriotas e amigos que os conheci durante essa jornada, pelo acolhimento, troca de experiências e ajuda na minha integração.

À todos vocês, o meu muito *khanimabo* (obrigado).

Palavras-chave

Meta-avaliação, Exercícios Parametrizados, Teoria Clássica de Testes, Teoria de Resposta ao Item

Resumo

Testes constituídos de questões de escolha múltipla são diversamente utilizados por professores na avaliação dos seus alunos. Porém, os resultados desses testes dizem mais do que a classificação final do aluno: aprovado ou reprovado. Neste estudo, recorrendo à Teoria Clássica de Teste (TCT) e à Teoria de Resposta ao Item (TRI), avaliamos testes constituídos por questões de escolha múltipla, considerados na unidade curricular de Bioestatística. Os objetivos foram: (i) verificar, por um lado, a qualidade global do conjunto dos 5 mini testes aplicados e, por outro lado, dos 37 itens individualmente; e (ii) analisar o nível de habilidade dos 111 alunos avaliados. As estatísticas produzidas pela TCT permitiram concluir que o conjunto dos 5 mini testes tinha boa consistência interna (com possibilidades de aumentar, eliminando uma questão tida como má) e que era composto por itens com diferentes níveis de dificuldade (28,83% - 95,50%) e de discriminação (0,0357 - 0,6586).

Três modelos da TRI foram testados sendo o modelo logístico de dois parâmetros (2PL) o que apresentou melhor ajuste. Os parâmetros do modelo 2PL foram estimados pelo método de estimação da máxima verossimilhança marginal. As curvas características dos itens contribuíram para a identificação de questões com diferentes níveis de dificuldade e de discriminação. As curvas de informação de itens apresentaram o nível de contribuição de cada questão na estimação das habilidades dos alunos e, a curva de informação de teste mostrou que a discriminação máxima é atingida para alunos abaixo da média na escala de habilidade, isto é, o teste mostrou ser mais eficaz para estudantes com nível de habilidade baixa a média.

Espera-se que este estudo demonstre a importância de avaliar as avaliações e motive a elaboração de testes com questões cujas características se adequem aos níveis de habilidades dos alunos, sem nunca esquecer contudo, os objetivos educacionais e as competências a fornecer. A parametrização de exercícios corresponde a uma técnica que permite considerar o mesmo conteúdo curricular sem usar, taxativamente, o mesmo enunciado. Neste trabalho ilustramos alguns exercícios parametrizados que podem ser incorporados num banco de questões.

Keywords

Meta-evaluation, Parameterized Exercises, Classical Tests Theory, Item Response Theory

Abstract

Tests consist of multiple choice questions are variously used by teachers in assessing their students. However, the results of these tests tell more than the final grade of the student: pass or fail. In this study, using Classical Test Theory (CTT) and Item Response Theory (IRT), we evaluate tests consisting of multiple choice questions, considered in the course of Biostatistics. The objectives were: (i) to determine, on one hand, the overall quality of all given five mini-tests, and, on the other hand, the 37 individual items; and (ii) to analyse the ability level of the 111 assessed students. The statistics produced by TCT showed that the set of 5 mini-tests had good internal consistency (with the possibility of increasing, eliminating a question seen as bad) and was composed of items with different levels of difficulty (28.83% to 95 50%) and discrimination (0.0357 to 0.6586).

Three IRT models were tested and the two-parameter logistic model (2PL) presented the best fit. The 2PL model' parameters were estimated by the method of marginal maximum likelihood estimation. The items characteristic curves contributed to identify questions with different levels of difficulty and discrimination. The items information curves showed the level of contribution of each question to the estimation of students' ability and the test information curve showed that the maximum discrimination is achieved for students below average in ability's scale, that is, the test was more effective for students with ability level between low and medium.

It is hoped that this study demonstrates the importance of assessing the evaluations and motivate the development of tests with questions whose characteristics suited to students' ability levels, never forgetting, however, the educational objectives and skills to provide. The parameterization exercises correspond to a technique that allows to consider the same curriculum content without using, exclusively, the same statement. In this paper we illustrate some parameterized exercises which can be incorporated into a bank questions.

Conteúdo

Conteúdo	i
Lista de Figuras	iii
Lista de Tabelas	v
Abreviaturas	vii
1 Introdução	1
1.1 Antecedentes e motivação	1
1.2 Objetivo	2
1.3 Pertinência	4
1.4 Organização	5
2 Modelos Matemáticos	7
2.1 Teoria Clássica de Teste	8
2.2 Teoria de Resposta ao Item	15
2.3 Modelos da TRI	19
2.3.1 Modelo logístico de 1 parâmetro (1PL)	21
2.3.2 Modelo logístico de 2 parâmetros (2PL)	22
2.3.3 Modelo logístico de 3 parâmetros (3PL)	23
2.3.4 Função de informação do item e função de informação do teste	24
2.4 Estimação dos Parâmetros	25
2.4.1 Estimação dos parâmetros dos itens sendo conhecidas as habilidades .	26
2.4.2 Estimação das habilidades sendo conhecidos os parâmetros dos itens .	29
2.4.3 Estimação dos parâmetros dos itens e das habilidades	31

2.4.4	Abordagem de Bock & Lieberman	32
2.4.5	Abordagem de Bock & Aitkin	35
2.4.6	Algoritmo EM	37
3	Aplicação	43
3.1	Descrição dos dados1 via TCT	44
3.2	Análise dos dados1 via TRI	46
3.3	Análise dos dados2	53
4	Exercícios parametrizados	59
4.1	Conteúdos tratados	60
4.2	Descrição e proposta de resolução de exercícios	62
4.2.1	Exercício 1	63
4.2.2	Exercício 2	65
4.2.3	Exercício 3	67
4.2.4	Exercício 4	70
	Considerações Finais	73
	Bibliografia	77
	Apêndices	83
	Anexo I – Programa de Bioestatística	103
	Anexo II – Teste aplicado aos alunos	107

Lista de Figuras

2.1	Exemplo de um curva característica de item	17
2.2	Curvas características de itens	18
3.1	Índices de dificuldade e de discriminação do teste	46
3.2	Curvas características dos 37 itens do teste	49
3.3	Curvas de informação dos 37 itens do teste	50
3.4	Comparação de duas CIIs do teste	51
3.5	Curva de informação do teste	52
3.6	Distribuição dos alunos na escala das habilidades	52

Lista de Tabelas

2.1.1	Recomendações sobre a confiabilidade do teste	11
2.1.2	Recomendações sobre a qualificação do índice de dificuldade dos itens	12
2.1.3	Recomendações sobre a qualificação do índice de discriminação dos itens . .	13
3.1.1	Resumo descritivo dos dados1 via TCT	45
3.2.1	Comparação do ajustamento dos modelos 1PL e 2PL	47
3.2.2	Comparação do ajustamento dos modelos 1PL e 3PL	47
3.2.3	Comparação do ajustamento dos modelos 2PL e 3PL	48
3.3.1	Nível de concordância	54
3.3.2	Nível de concordância para cada item	55
3.3.3	Nível de concordância para cada momento de avaliação	55
3.3.4	Itens mais fáceis e mais difíceis na percepção dos avaliadores	56
3.3.5	Itens mais fáceis e mais difíceis para os alunos	57
A.1	Índices de dificuldade, discriminação e de consistência interna de cada item obtidos via TCT	83
A.2	Parâmetros de dificuldade, discriminação e acerto casual de cada item esti- mados via TRI pelos modelos 1PL, 2PL e 3PL	85
A.3	Habilidade de cada aluno estimada via TRI pelos modelos 1PL, 2PL e 3PL .	87

Abreviaturas

CCI Curva Característica do Item

CII Curva de Informação do Item

CIT Curva de Informação do Teste

DMat Departamento de Matemática

EMVC Estimação da Máxima Verosimilhança Conjunta

EMVM Estimação da Máxima Verosimilhança Marginal

PmatE Projeto Matemática Ensino

SEM Erro Padrão de Medição

TCT Teoria Clássica de Testes

TRI Teoria de Resposta ao Item

W Coeficiente de concordância W de Kendall

1PL Modelo logístico de 1 parâmetro

2PL Modelo logístico de 2 parâmetros

3PL Modelo logístico de 3 parâmetros

Capítulo 1

Introdução

“The mind that opens to a new idea never returns to its original size”

Albert Einstein (1879 - 1955)

1.1 Antecedentes e motivação

A educação é um processo dinâmico e contínuo que acompanha o indivíduo em todas as facetas da vida. Por ser um processo de extrema importância, a sociedade, desde a camada mais baixa, e os governos, ao mais alto nível, apostam na educação como uma via através da qual o cidadão possa garantir a sua integração com os demais, evoluir e fazer aplicações dos seus conhecimentos em vários domínios do saber. Na educação escolar, o processo educativo é regido por normas, objetivos concretos e padrões com a finalidade de se alcançar o conhecimento científico. Alguns dos instrumentos habitualmente utilizados, para verificar se o conhecimento científico é alcançado e até que nível ele é alcançado, são as provas avaliativas ou exames (doravante, simplesmente, designados por testes). Concretamente, no processo de ensino e no processo de aprendizagem, os testes são utilizados para verificar se os objetivos programáticos para um determinado conteúdo curricular, unidade temática, capítulo, programa ou ciclo acadêmico foram, efetivamente, alcançados. Espera-se que os resultados obtidos em tais testes sejam o reflexo do desempenho dos examinandos, os quais poderão ditar o seu futuro em termos de distinção e progressão. Se no processo do ensino se tem, fundamentalmente, os educadores (professores ou entidades) que elaboram e aplicam os testes em função dos objetivos que pretendem alcançar então, no processo de aprendizagem, têm-se os examinandos (alunos) que devem possuir determinadas capacidades (habilidades)

para responderem positivamente aos testes.

Ao nível da Universidade de Aveiro, e especificamente no Departamento de Matemática (DMat) local onde foi desenvolvido este trabalho, as avaliações acompanham o processo de ensino e o processo de aprendizagem. A unidade curricular de Bioestatística, da responsabilidade do DMat para os alunos do 1^o ano do curso de Biologia e do 2^o do curso de Biologia e Geologia, foi escolhida para estudo nesta dissertação pois, pela primeira vez, no ano lectivo 2015/2016 mereceu destaque face ao interesse dos professores em reformular os modelos de avaliação. As avaliações foram parceladas em dois modelos: testes escritos, com 75% da nota final, e um trabalho em grupo, com os restantes 25%. Os testes consistiam de questões de escolha múltipla com apenas uma opção correta cada, permitindo que as respostas fossem dicotomizadas e, conseqüentemente, corrigidas no critério de certo ou errado. Anteriormente, embora algumas avaliações incluíssem questões de escolha múltipla seleccionadas a partir do moodle, o peso destas avaliações sob a nota final era de apenas 5%. O objetivo, a médio prazo, será de criar um banco de dados com perguntas parametrizadas para diferentes conteúdos e, eventualmente no futuro, com a implementação do mesmo modelo noutras unidades curriculares, comparar o desempenho dos alunos de diversos cursos ao longo de anos. Com base na mudança de paradigma nos critérios de avaliação e no propósito dos professores desta unidade curricular, vimos uma oportunidade de desenvolver um estudo que contribuísse com a efetivação de tal propósito.

1.2 Objetivo

Outrora as avaliações tinham como finalidade a obtenção de resultados classificatórios. Hoje, com os investimentos feitos no processo educativo, interesses há não só em fazer estudos comparativos dos examinandos em diferentes épocas ou dos examinandos de diferentes grupos mas também em saber o que está por detrás desses resultados (por exemplo, as habilidades cognitivas dos examinandos e a qualidade dos instrumentos de avaliação). Para lograr esses intentos há necessidade de avaliar as avaliações, isto é, verificar o nível de qualidade da própria avaliação com base em critérios adequados. Este processo é fundamental pois

permite verificar se os resultados gerados pela avaliação são úteis e contribui para obtenção de informações atinentes às limitações e potencialidades da avaliação. A avaliação da avaliação é conhecida na literatura como meta-avaliação (Scriven, 1969). Etimologicamente meta significa “depois de”, contudo, a meta-avaliação pode ser formativa (realizada continuamente visando a melhoria do processo avaliativo) quanto sumativa (realizada no fim do processo avaliativo examinando cuidadosamente os resultados com a finalidade de analisar o seu valor e detetar pontos fracos e fortes das avaliações anteriores). Para Severo & Tavares (2010), o controle da qualidade dos teste escritos são de extrema importância e a meta-avaliação permite determinar se os objetivos definidos para uma dada disciplina estão sendo alcançados e avaliados através do teste proposto. Independentemente de ser ou não formativa, a meta-avaliação deve-se orientar por alguns critérios consensuais entre vários autores: a validade (interna e externa), confiabilidade, objetividade, credibilidade, utilidade ou importância, custo-benefício, relevância, abrangência, oportunidade e dissiminação (Scriven, 1991; Davidson, 2005; Stufflebeam, 2011).

A Teoria Clássica de Testes (TCT) e a Teoria de Resposta ao Item (TRI), abordadas nesta dissertação, possibilitam a partir dos resultados obtidos nas avaliações, respetivamente, a obtenção de estatísticas que permitem avaliar a qualidade de um teste e a definição de modelos para medir as habilidades dos examinandos e estimar os níveis de dificuldades, de discriminação e de acerto casual presentes nas perguntas. Os modelos da TRI estabelecem relação entre a probabilidade de um indivíduo responder corretamente a uma pergunta e as suas habilidades. Em particular, em contexto educativo, os modelos da TRI são utilizados para modelar a probabilidade de um aluno responder corretamente a uma questão de um teste em termos da capacidade do indivíduo e das características da questão.

Nesta dissertação pretendemos fazer uma meta-avaliação das questões de escolha múltipla contidas nos testes de Bioestatística do último ano letivo 2015/2016. Mais ainda, com vista a experimentar as dificuldades inerentes à construção de potenciais questões com enunciados não estáticos (variáveis cada vez que se recorre a ela), na presente dissertação contribuimos ainda com um conjunto de questões parametrizadas que podem ser incorporadas num banco de questões de Estatística.

Fruto deste trabalho de meta-avaliação e parametrização de exercícios, outras questões, como as que se seguem, foram despertando o nosso interesse: como os professores podem elaborar ou encontrar questões adequadas ao tipo de alunos que têm? Até que ponto questões parametrizadas podem ajudar os professores a minimizar a tarefa de elaborar periodicamente questões para avaliar os mesmos conteúdos? Como os professores podem elaborar ou encontrar questões que garantem confiabilidade e validade dos resultados? Será que a forma como os professores encaram um teste é a mesma com que os alunos o encaram?

1.3 Pertinência

A TCT foi reconhecida depois da divulgação do trabalho de Gulliksen (1950) e a TRI, depois da divulgação do trabalho de Lord & Novick's (1968). Estas teorias têm suas origens na Psicometria. Embora não sejam teorias novas, a sua difusão e aplicação não são de domínio de muitos investigadores. A realização deste estudo, com recurso a TCT e TRI, constitui um contributo para os profissionais ligados ao processo educativo. Áreas como pesquisa de mercado, pesquisa de marketing, pesquisas psicológicas, satisfação do cliente, produção de indicadores sócio-económicos, só para destacar algumas, podem aplicar estas teorias. Efetivamente, quanto mais estatísticos, ou outros profissionais que usam a Estatística nas suas pesquisas, souberem da aplicabilidade destas teorias, maiores serão as hipóteses de entenderem melhor os problemas associados à avaliação e, conseqüentemente, obterem resultados satisfatórios, consistentes e válidos.

As avaliações devem satisfazer às reais intenções do avaliador e corresponder às expectativas dos avaliados. Nesse contexto, a avaliação tornar-se-á num fator motivador para a aprendizagem, um meio de distinguir os mais empenhados e direcionar melhor os indivíduos às áreas de seu interesse e domínio próprio, num instrumento que facilite a atividade docente. A TCT e TRI são teorias indicadas para suprir essas necessidades e, como mencionado acima, estas teorias e suas aplicações serão abordadas neste trabalho.

1.4 Organização

Esta dissertação, cujo estudo focaliza a aplicação da TCT e dos modelos unidimensionais da TRI para dados dicotômicos, está estruturada em mais três capítulos para além deste primeiro e foi usada a plataforma *SageMathCloud* com recurso à linguagem \LaTeX . O segundo capítulo ficou reservado à abordagem da TCT e da TRI. Em relação à TRI, para além da descrição dos respetivos modelos, são apresentadas as abordagens para estimação das habilidades dos examinandos e dos parâmetros dos itens. O terceiro faz a aplicação da TCT e da TRI aos dados reais, relativos às avaliações na unidade curricular de Bioestatística realizadas em 2015/16 na Universidade de Aveiro, bem como a discussão dos resultados. Finalmente, tem-se o quarto capítulo onde são apresentados, a título modelo, quatro exercícios parametrizados. Cada enunciado é acompanhado da respetiva descrição, proposta de resolução e breves comentários de alguns erros cometidos pelos alunos. Para epílogo da dissertação, algumas considerações finais, em jeito de conclusão, limitações e sugestões para trabalhos futuros, são apresentadas.

Capítulo 2

Modelos Matemáticos

A investigação vai além de analisar características ou fenómenos observáveis. A psicometria, tida como ramo da Psicologia que consiste em técnicas de medição do comportamento do ser humano por meio de processos mentais, foi ao longo dos anos, e continua sendo bastante aplicada por psicólogos para medir, de forma adequada, traços latentes que melhor se desejam conhecer (Pasquali, 2009; Primi, 2012; Ureña et al , 2015). Embora a psicometria esteja estritamente ligada à Psicologia, sua aplicação é extensiva a outras áreas de conhecimento como, por exemplo, avaliação educacional, pesquisa do mercado, satisfação do cliente e produção de indicadores sócio-económicos. Na área da Educação e do Ensino em particular, pode interessar aos investigadores estudar traços latentes como a habilidade que os alunos têm numa determinada área de conhecimento, e analisar características das questões que compõem a avaliação (Severo & Tavares, 2010; Pereira, 2015; Costa, 2005). É evidente que traços como a habilidade matemática de um indivíduo, a tendência de compra de um produto, a atitude, outros, não são visíveis. Contudo, através de respostas que um indivíduo, ou grupo de indivíduos, dá a um instrumento de medição como, por exemplo, um teste composto por várias questões, é possível tirar conclusões de como o traço em causa se relaciona com cada questão do teste ou com todo o teste. A psicometria procura dar significado às respostas dadas pelos indivíduos a um conjunto de questões que lhes foram colocadas e, para tal, usa duas teorias conhecidas como Teoria Clássica de Testes (TCT) e Teoria de Resposta ao Item (TRI).

2.1 Teoria Clássica de Teste

A TCT vem sendo referenciada há bastante tempo. De acordo com Muñiz (2010), o modelo da TCT e os conceitos relacionados aos seus procedimentos são citados em obras de Spearman (1907, 1913) sendo que mais tarde, Gulliksen (1950) forneceu detalhes adicionais e axiomatizou a TCT. Este modelo é aditivo e diz que o resultado (X) observado no teste de um indivíduo resulta da adição do resultado verdadeiro (T) do indivíduo com o erro de medição (E) associado a esse mesmo resultado, isto é,

$$X = T + E. \quad (2.1)$$

Admitindo um erro de medição com média nula, o valor esperado para o resultado observado será o resultado verdadeiro, isto é, $E(X) = T$. A TCT tem como tarefa definir estratégias para controlar a magnitude dos erros resultantes de defeitos do próprio teste, fatores históricos e ambientais aleatórios, estereótipos do indivíduo, entre outros (Campbell & Stanley, 1973). Para além de assumir o erro de medição numa população de examinandos como nulo, Hambleton & Jones (1993) acrescentam que, o resultado verdadeiro e o erro correspondente não estão correlacionados e, ainda, que o erro dos resultados em testes paralelos não estão correlacionados.

As bases que sustentam a TCT deixam claro que para a obtenção do resultado verdadeiro, o teste deve medir exatamente o traço que se pretende avaliar no indivíduo, isto é, a avaliação do traço (habilidade matemática, por exemplo) de um indivíduo depende, fundamentalmente, da qualidade do teste. Para Carmines & Zeller (1987) e Primi (2012) dois fatores (propriedades métricas básicas) garantem a legitimidade de um teste: a validade e a confiabilidade. Na verdade, muitas decisões são tomadas em função da interpretação de resultados observados num teste ou instrumento de medição e, a partir dessa interpretação, as ideias ou teorias podem ser ou não confirmadas. Daí que Hogan & Agnello (2004), para além de considerarem a validade e a confiabilidade como pilares da qualidade psicométrica, fazem menção a casos em que vários artigos foram classificados como rejeitados ou como necessitando de mais revisão pela falta de validade e confiabilidade dos dados. Portanto, um teste ou instrumento de medição deve ser tanto confiável quanto válido.

Um instrumento de medição é considerado válido quando cumpre o objetivo para o qual foi proposto (Carmines & Zeller, 1987; Primi, 2012). Por outras palavras, espera-se que os resultados observados num teste devem estejam associados ao traço latente que o teste se propõe a medir. Por exemplo, se se pretende medir a temperatura de um indivíduo é justo que se utilize um termómetro e não uma fita métrica. Para o caso de um teste, é de se esperar que todos os itens que o compõe estejam fortemente relacionados com o traço que se pretende medir pois, caso contrário, os resultados não servirão para o fins previstos e consequentemente descartados. Ainda de acordo com Carmines & Zeller (1987) e Pasquali (2009), para que um teste seja considerado válido deve ter evidência de validade aparente (se o teste avalia aquilo que aparentemente pretende avaliar), validade do conteúdo (grau em que um teste evidencia a predominância do conteúdo que pretende avaliar), validade de critério (estabelece a validade do teste comparando-o com algum critério, isto é, o teste é utilizado para estimar um traço que é externo a ele) e validade do construto (grau em que um teste se relacione consistentemente com outros derivados da mesma teoria e conceitos que estão a ser avaliados).

A confiabilidade corresponde à consistência ou capacidade de reproduzir o desempenho de um indivíduo no teste. É de se esperar que, a partir de um teste com alta confiabilidade, se chegue às mesmas conclusões sobre o desempenho de um indivíduo, quando submetido ao mesmo teste por duas vezes, por exemplo (Carmines & Zeller, 1987; De Villes, 2012). Mais ainda, um teste com uma confiabilidade reduzida, pode resultar em resultados muito diferentes para o indivíduo entre as duas administrações de teste. Se o segundo caso se verificar, os resultados observados não seriam recomendados para medir o traço em avaliação. Para Hayes (1998), a confiabilidade é a medida em que as medições estão livres da variância dos erros aleatórios, isto é, o resultado medido deve refletir o resultado verdadeiro. Portanto, a confiabilidade pode ser deduzida, conforme mostra a expressão (2.2), como a razão entre a variância do resultado verdadeiro e a variância do resultado observado e, consequentemente, quanto mais próximo o resultado observado estiver do resultado verdadeiro, maior será a confiabilidade do teste:

$$Confiabilidade = \frac{Var(T)}{Var(X)} = \frac{\sigma_T^2}{\sigma_X^2}. \quad (2.2)$$

Note-se que o resultado verdadeiro não é conhecido e conseqüentemente a sua variância também, impossibilitando a estimação da confiabilidade. Contudo, Díaz et al (2003) e Primi (2012) depois de mostrarem que a confiabilidade é medida pela correlação entre os resultados observados identificaram alguns procedimentos, comumente utilizados, que permitem determinar a confiabilidade de um teste: (i) teste re-teste ou coeficiente de estabilidade - procura verificar se as pontuações, depois de o mesmo teste ser administrado duas vezes ao mesmo grupo, manter-se-ão estáveis resultando numa correlação positiva muito forte; (ii) formas paralelas ou coeficiente de equivalência - procura verificar se as duas formas de testes (similares em conteúdos e instruções) administradas ao mesmo grupo são equivalentes na expectativa dos padrões de respostas serem muito semelhantes resultando numa correlação positiva forte; (iii) metades partidas (*split-half*) - uma vez o teste dividido, aleatoriamente, em duas partes equivalentes é administrado e verifica se os resultados dos respondentes proporcionam uma correlação positiva forte; (iv) covariância entre os itens do teste ou consistência interna - procura verificar até que ponto cada item mede o traço em consideração e o que se espera é que a correlação entre resposta a cada item seja forte e positiva com a pontuação total. Este coeficiente de confiabilidade (consistência interna) pode ser determinado através das fórmulas de

$$\text{Alpha de Cronbach } (\alpha) : \quad \alpha = \frac{n}{n-1} \left[1 - \frac{\sum_{i=1}^n s_i^2}{s_n^2} \right] \quad \text{e,} \quad (2.3)$$

$$\text{Kuder-Richardson } (KR_{20}) : \quad KR_{20} = \frac{n}{n-1} \left[1 - \frac{\sum_{i=1}^n p_i (1-p_i)}{s_n^2} \right] \quad (2.4)$$

onde $i = 1, 2, \dots, n$ são os itens, s_i^2 é a variância da pontuação de cada item, s_n^2 é a variância da pontuação total e p_i é o índice de dificuldade do item.

O coeficiente de alpha de Cronbach, desenvolvido por Cronbach em 1951, é uma generalização da fórmula KR_{20} , proposto por Kuder-Richardson em 1937 para itens dicotômicos, e visa medir a consistência interna de um instrumento, isto é, avaliar a magnitude em que os itens estão correlacionados. É considerado uma das ferramentas estatísticas mais importantes e difundidas em pesquisas que envolvem construção de teste e sua aplicação (Cortina, 1993). Este valor varia entre 0 e 1, podendo assumir valores negativos quando a correlação entre

os itens for negativa. Nesse caso, os pontos dos itens devem ser re-codificados de forma a assegurar que todos os itens estão codificados na mesma direção conceptual (Maroco & Marques, 2006). Embora não exista consenso quanto ao valor do alpha de Cronbach para uma boa confiabilidade do instrumento, o valor mínimo aceitável é de 0.7 podendo ser menos, dependendo do número de itens que compõem o instrumento ou do número de indivíduos que respondem. A Tabela 2.1.1, adaptada por Peterson (1994), apresenta recomendações de vários autores relativamente ao nível de confiabilidade:

Autor	Situação	Nível de α
Davis (1964, p.24)	Previsão individual	Acima de 0.75
	Previsão para grupos de 25 - 50 indivíduos	0.5
	Previsão para grupos acima de 50 indivíduos	Abaixo de 0.5
Kaplan & Sacuzzo (1982, p.106)	Investigação básica	0.7 - 0.8
	Investigação aplicada	0.95
Murphy & Davidsholder (1988, p.89)	Nível inaceitável	< 0.6
	Nível baixo	0.7
	Nível moderado a alto	0.8 - 0.9
	Nível alto	> 0.9
Nunnally (1967, p.245-246)	Investigação preliminar	0.7
	Investigação básica	0.8
	Investigação aplicada	0.9 - 0.95

Tabela 2.1.1: Recomendações sobre a confiabilidade do teste

A confiabilidade mais do que fornecer um índice para avaliar a qualidade de um teste, possibilita a estimação do erro padrão de medição (SEM - *Standard Error of Measurement*). O SEM pode ser utilizado para construir o intervalo de confiança onde se espera que se encontre o resultado verdadeiro do indivíduo. O valor do SEM é calculado através da fórmula:

$$SEM = \sigma_X \sqrt{1 - r}, \quad (2.5)$$

onde σ_X é o desvio padrão do teste e r é a estimativa da confiabilidade. A interpretação do SEM baseia-se numa distribuição normal.

No contexto da TCT, uma vez construído o teste e aplicado ao grupo alvo, algumas estatísticas, tais como índice de dificuldades e índice de discriminação, são necessárias para verificar as propriedades do mesmo.

O índice de dificuldades é a proporção dos examinandos que responderam corretamente ao item. Trata-se de um valor que varia entre 0 (nenhum dos examinandos respondeu corretamente ao item) e 1 (todos os examinandos responderam corretamente ao item). Obviamente, os itens com índice de dificuldades próximos de 0 são tidos como difíceis e, os próximos de 1, como fáceis. O ideal seria usar um teste composto por itens fáceis, médios e difíceis por forma a evidenciar as diferenças individuais dos examinandos. Contudo, há que ter muita atenção em relação aos itens com baixo valor no índice de dificuldade pois, fatores tais como exigência demais do item em relação às habilidades dos examinandos, itens com mais de uma resposta correta, ambiguidade do item e inclusive o uso de uma chave de resposta errada podem estar diretamente associados. A seguir são apresentados, por meio da Tabela (2.1.2), os valores recomendados para avaliar o índice de dificuldade (Tavakol & Dennick, 2011):

Índice de dificuldade	Pergunta (Item)
[0.00; 0.30]	Difícil
]0.30; 0.80]	Dificuldade média
]0.80; 1.00]	Fácil

Tabela 2.1.2: Recomendações sobre a qualificação do índice de dificuldade dos itens

O índice de discriminação é usado para discriminar os examinandos com nível alto de conhecimentos dos examinandos com nível baixo. Para o efeito é usado o coeficiente de correlação ponto bisserial ($\hat{\rho}_{pb}$) para quantificar o grau associação entre a pontuação (dicotómica - "1" se

acertar e "0" se não acertar) de cada item e a pontuação total do teste. Assim, se os examinandos com maior nível de conhecimento responderem corretamente ao item em relação aos que têm um nível menor, a correlação ponto bisserial será positiva e, negativa, caso contrário. O valor do índice de discriminação assim definido, varia entre -1 e 1 mas os desejáveis são os positivos e é dado por

$$\hat{\rho}_{pb} = \frac{\bar{X}_p - \bar{X}_t}{s_t} \sqrt{\frac{p}{1-p}}, \quad (2.6)$$

onde:

\bar{X}_p - média das pontuações dos examinandos que acertaram o item;

\bar{X}_t - média total das pontuações;

s_t - desvio padrão das pontuações no teste de todos os examinandos;

p - proporção dos examinandos que acertaram o item.

Por forma a avaliar o índice de discriminação, a Tabela (2.1.3), adaptada por Ebel (1979), avança com algumas recomendações:

Índice de discriminação	Discriminação
[-1.00; -0.20]	Negativa
]-0.20; 0.19]	Fraca
]0.19; 0.29]	Suficiente
]0.29; 0.39]	Boa
]0.39; 1.00]	Muito boa

Tabela 2.1.3: Recomendações sobre a qualificação do índice de discriminação dos itens

Existe uma forte ligação entre o índice de discriminação, dado pelo coeficiente de correlação ponto bisserial, e o índice de dificuldades. Na verdade, os itens com índices de dificuldade de 0 ou 1 têm, sempre, discriminação nula, mas os itens com índices de dificuldade muito próximos de 0.5 têm alto poder discriminativo.

A TCT foi, outrora, bastante utilizada para desenvolvimento de testes psicológicos e o mérito atual não lhe é tirado pois continua sendo utilizada por psicólogos e investigadores devido a sua facilidade em analisar e interpretar os resultados. Esta teoria dá mais enfoque ao resultado de todo o teste e, portanto, todas as análises e interpretações são sempre associadas como um todo. A aplicabilidade plena da TCT foi questionada nos trabalhos de Lord & Novick's (1968) que, para além de discutir algumas limitações da TCT, fazem menção à TRI como a solução dessas limitações. Desde então, as bases das teorias da medida psicológica mudaram significativamente em virtude da eficácia da TRI em resolver problemas práticos de testes (Hambleton, 1982). Primeiramente, foram desenvolvidos modelos para itens dicotômicos ou itens de escolha múltipla (com resultado certo ou errado), (Rasch, 1960) e anos depois foram desenvolvidos modelos para itens politômicos, caso de modelos nominais e modelos ordinais com a escala de Likert, por exemplo, modelo de escala gradual (Andrich, 1978), modelo de crédito parcial (Masters, 1982) e modelo de crédito parcial generalizado (Muraki, 1992). A maioria destes modelos têm como um dos pressupostos a unidimensionalidade do traço latente. Porém, estudos indicam a extensão para modelos multidimensionais (Linden & Hambleton, 1997).

Embreston (1996), Hambleton (1997) e Fan (1998) apresentam uma série limitações da TCT superadas pela TRI:

- Requer um número grande de participantes pois todas as medidas dependem da amostra dos indivíduos que respondem ao teste;
- Se o mesmo traço é medido por testes diferentes, os resultados não são expressos na mesma escala de medida, impedindo uma comparação direta.
- O erro padrão é o mesmo para o instrumento como um todo;
- Os parâmetros do item dependem da amostra na qual o instrumento foi aplicado;
- A confiabilidade é definida em termos de testes paralelos (testes com pontuações iguais e variâncias iguais);
- A discriminação do item é baseado na pontuação total do teste;

- O número de itens influencia na confiabilidade do teste.

Segundo Andrade et al (2000) as limitações da TCT, apresentadas acima, são fortemente influenciadas pela sua característica básica em que resultados encontrados dependem de um conjunto particular de itens que compõem o teste, ou seja, todas as análises e interpretações estão sempre associadas ao teste como um todo. Numa abordagem em relação a TRI, Andrade et al (2000) indicam algumas aplicações possíveis:

- Comparação entre grupos de examinandos submetidos ao mesmo teste ou a teste com alguns itens em comum;
- Comparação entre examinandos de um mesmo grupo submetidos a teste único, parcialmente ou totalmente diferente;
- Acompanhar a evolução de um grupo de examinandos ao longo de um período;
- Comparar desempenho entre diferentes instituições (por exemplo, entre escolas privadas e públicas);
- Avaliar corretamente os itens independentemente das habilidades dos examinandos, ou as habilidades independentemente dos parâmetros dos itens.

Os avanços conseguidos na TRI são o complemento daquilo que a TCT tem como limitação e não constituem substituição em si (Araújo et al, 2009). Dado ao contributo da TRI nos estudos apontados por Andrade et al (2000), na avaliação e no aperfeiçoamento de testes, são abordados nas Secções 3.2 a 3.4 o conceito, os pressupostos, os modelos da TRI e a estimação dos parâmetros dos itens e das habilidades dos alunos.

2.2 Teoria de Resposta ao Item

Baker (1992) faz menção a vários estudos que visaram o desenvolvimento e consolidação da TRI mas foram os trabalhos de Lord (1952) que permitiram apresentar um modelo teórico e os métodos para estimar os parâmetros dos itens utilizando o modelo da ogiva normal. Num estudo paralelo, Rasch (1960) deu seu contributo a TRI ao desenvolver modelos de um

parâmetro, ordinal e multivariado sendo o primeiro, o modelo mais conhecido e aplicado. Finalmente, Birnbaum (1968) substituiu o modelo de ogiva normal de dois parâmetros proposto por Lord pelo modelo logístico de dois parâmetros e ainda acrescentou, ao modelo, o parâmetro do acerto ao acaso que ficou conhecido como modelo logístico de 3 parâmetros.

A TRI é um conjunto de modelos matemáticos e estatísticos definidos com intuito de explicar a resposta de um indivíduo a um item. Estes itens podem medir diferentes traços latentes de um indivíduo tais como as habilidades, as tendências, traços comportamentais, atitude, proficiência num teste ou questionário (Embreston, 1996; Hambleton, Swaminathan & Roger, 1991). Por outras palavras, a partir das respostas dadas por um indivíduo pode-se inferir sobre o traço latente, estabelecendo possíveis relações entre as respostas observadas deste indivíduo com o nível do seu traço latente. Estas relações podem ser expressas por meio de uma equação matemática que descreve a forma de função que estas relações assumem. Concretamente, a TRI pressupõe que a probabilidade do indivíduo j (casualmente seleccionado de uma amostra ou população de J indivíduos), com habilidade θ_j , responder corretamente a um item qualquer i (num conjunto de n questões) com certos parâmetros (parâmetro de dificuldade b_i , parâmetro de discriminação a_i e parâmetro de acerto casual c_i)¹, depende da habilidade do indivíduo e dos parâmetros do item. O parâmetro b_i representa a localização de um item na mesma escala da habilidade. O parâmetro a_i discrimina os indivíduos mais proficientes dos menos proficientes e permite, em indivíduos com mesma habilidade, distinguir qual o item que melhor os discrimina. O parâmetro c_i indica a probabilidade de um indivíduo com baixa habilidade responder corretamente o item.

Teoricamente, é de se esperar que um indivíduo com habilidade elevada, numa área de conhecimento, tenha mais probabilidade de responder corretamente ao item. Esta probabilidade é melhor apresentada através da curva característica do item (CCI) cujas expressões matemáticas, dependendo dos parâmetros do item, são apresentadas nas Subsecções 2.3.1 a 2.3.3. Portanto, para um dado item, a CCI descreve como a probabilidade de acerto a esse item

¹Por simplificação de linguagem escreveremos “dificuldade” para se referir ao parâmetro de dificuldade, “discriminação” para se referir ao parâmetro de discriminação e “acerto casual” para se referir ao parâmetro de acerto casual.

varia com a variação da habilidade, desde que sejam conhecidos os parâmetros do item (ver Figura 2.1).

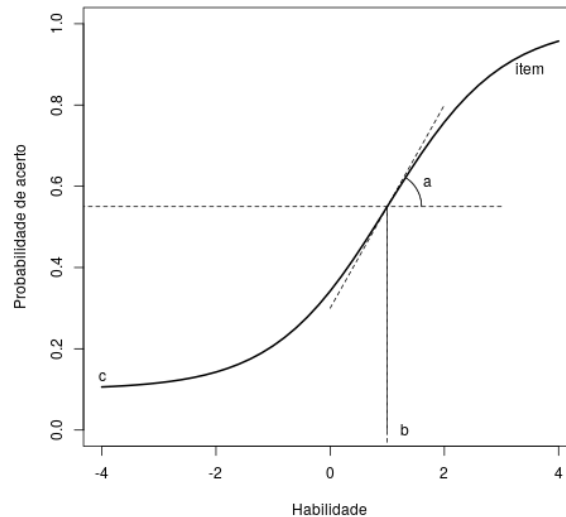
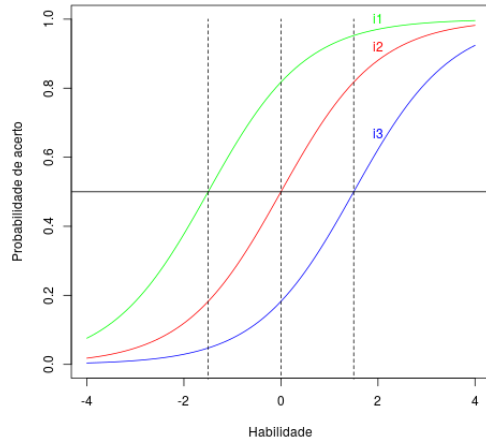


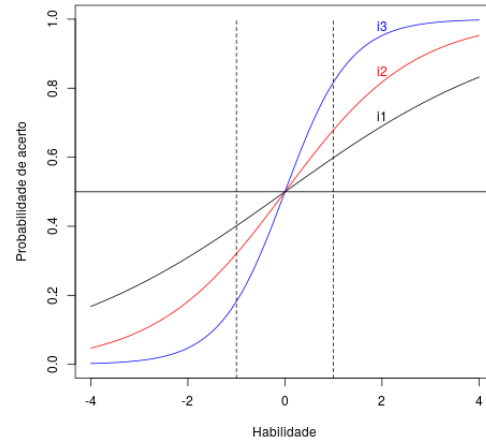
Figura 2.1: Exemplo de uma curva característica de item

É evidente que as curvas características dos itens (CCIs) serão diferentes para diferentes itens. Veja as situações (a) e (b) apresentadas na Figura (2.2). Na situação (a) é destacado o comportamento de três CCIs todas com discriminação igual a 1, acerto ao acaso nulo e parâmetros de dificuldades diferentes (-1.5, 0, 1.5). Observa-se que, para a mesma probabilidade de acerto de 50%, o item 3 requer mais habilidade em relação aos itens 1 e 2. Ademais, a probabilidade de acerto para o item 1 é sempre maior que os itens 2 e 3, para qualquer nível de habilidade. Nessas condições, o item 1 é tido como o mais fácil e o item 3 como o mais difícil. Na situação (b) são, também, apresentadas três CCIs mas desta vez, construídas tomando o mesmo nível de dificuldade (0), acerto ao acaso nulo e discriminações diferentes (0.4, 0.75, 1.5). Note que à medida que se aumenta o valor de discriminação, a CCI torna-se mais íngreme em torno de zero. Outra constatação é que analisando, por exemplo, a variação do valor da habilidade de -1.0 a 1.0, a probabilidade de responder corretamente ao item possui um suporte de variação com maior amplitude comparando aos suportes de variação para os restantes itens, aumentando a amplitude de cerca de 0.2 no item 1 para aproximadamente 0.62 no item 3. Pode-se afirmar que os itens que apresentam valores altos

do parâmetro de discriminação são itens cujas CCI's têm maior inclinação (curvas muito íngremes). Consequentemente, o item 3 discrimina melhor os indivíduos com habilidade em torno de 0 em relação aos itens 1 e 2.



(a) Com três parâmetros de dificuldade diferentes e os restantes índices fixos



(b) Com três parâmetros de discriminação diferentes e os restantes índices fixos

Figura 2.2: Curvas características de itens

O ideal é que sejam dados ou conhecidos quer os parâmetros dos itens quer as habilidades dos indivíduos. Contudo, nem sempre se tem estes valores havendo necessidade de estimá-los. A TRI apresenta modelos que podem ser usados para estimar esses valores desde que se tenha um instrumento para medir (teste) a habilidade. Embreston & Reise (2000) apresentam dois pressupostos para aplicação dos modelos da TRI para itens dicotômicos:

- Unidimensionalidade: os itens que compõem o teste devem medir apenas um único traço latente;
- Independência local: a probabilidade de responder a um item é determinada pela habilidade do aluno e não por suas respostas aos restantes itens, isto é, para uma determinada habilidade, as respostas aos diferentes itens do teste são independentes. Para Lord (1980) e Hambleton et al (1991) a independência local é uma implicação da unidimensionalidade.

2.3 Modelos da TRI

Neste trabalho são apresentados apenas os modelos logísticos para itens dicotômicos. Na dicotomização da variável resposta (Y) é comum atribuir o código “1” ao acontecimento associado sucesso ($Y = 1$) e o código “0” ao acontecimento insucesso ($Y = 0$). Assim, as probabilidades de ocorrência de sucesso e insucesso são, respetivamente, $P_i(X) = P(Y = 1)$ e $Q_i(X) = 1 - P_i(X) = P(Y = 0)$. Numa perspetiva de regressão, pode-se admitir que $P_i(X)$ está relacionada com uma ou mais variáveis preditoras X_i mas não se pode assumir que a relação seja linear da forma

$$P_i(X) = B_0 + B_1X_1 + \dots + B_kX_k, \quad i = 1, 2, \dots, n. \quad (2.7)$$

A variável dependente Y , como é óbvio, não é contínua mas o objetivo da regressão logística é prever $P(Y = 1)$ dado X , isto é, $P(Y = 1|X = x)$. Se a probabilidade condicional é a esperança condicional da variável de previsão, isto é, $P(Y = 1|X = x) = E[Y|X = x]$ pode-se afirmar que $P(Y = 1) = E(Y)$. O que se pretende são as previsões das probabilidades em função dos valores de X . Portanto, pode-se recorrer ao modelo de probabilidade linear $P_i(X) = B_0 + B_1X$ cujos parâmetros podem ser estimados através dos mínimos quadrados ordinários, lembrando que $P(Y = 1) = P_i(X)$. O problema com os mínimos quadrados ordinários é que a probabilidade $P_i(X)$ toma os valores entre 0 e 1 e o preditor linear $B_0 + B_1X$ pode tomar qualquer valor real e portanto não haverá garantias de que os valores previstos variem no intervalo correto a menos que sejam impostas fortes restrições nos coeficientes. Propõe-se a análise da distribuição de Y_i para mostrar que a previsão de $P_i(X)$ não pode ser feita a partir de um modelo linear.

Como

$$Y_i = \begin{cases} 1, & \text{se o evento ocorre com sucesso} \\ 0, & \text{caso contrário} \end{cases}$$

e ainda,

$$\begin{cases} Y_i = 1 & \rightarrow & P(Y_i = 1) = P_i(X) \\ Y_i = 0 & \rightarrow & P(Y_i = 0) = Q_i(X) \end{cases}$$

pode-se dizer que Y_i tem distribuição de Bernoulli com parâmetro $P_i(X)$. Assim,

$$Y_i \sim \text{Bern}[P_i(X)] \Rightarrow \begin{cases} E(Y_i) = P_i(x) \\ \text{Var}(Y_i) = P_i(x)Q_i(x) \end{cases}, \quad i = 1, 2, \dots, n.$$

Note-se que tanto a esperança como a variância dependem da probabilidade $P_i(X)$ o que significa que qualquer fator que afete a esta probabilidade irá alterar estas duas medidas. Portanto, um modelo linear em que as variáveis preditoras afetam a média mas assume que a variância é constante não é apropriado para analisar dados dicotômicos.

Para solucionar o problema acima referenciado deve-se encontrar uma função $h[P_i(X)] : h \in]-\infty, +\infty[$ que possa ser modelada pela função linear $B_0 + B_1X$. Na perspectiva de modelos lineares generalizados, a função h é designada por função de ligação (Nelder & Wedderburn, 1972). Assim, propõe-se a transformação logística que consiste em (i) transformar a probabilidade $P_i(X)$ em *odds* e (ii) definir uma nova medida designada por *logit* (logaritmo natural das *odds*). Se as *odds* para a ocorrência de um evento é o rácio entre a probabilidade de ocorrer $P_i(X)$ e de não ocorrer $Q_i(x)$ então o seu valor $\frac{P_i(X)}{Q_i(X)}$ é sempre positivo. O cálculo de *logit* garante que as probabilidades variem entre zero e um em todo o domínio real pois, se por um lado, à medida que a probabilidade aproxima-se de zero, as *odds* tendem para zero e o *logit* tende para $-\infty$, por outro lado, à medida que a probabilidade aproxima-se de um tanto as *odds* como o *logit* tendem para $+\infty$. Ademais, quando a probabilidade é 0.5, as respectivas *odds* tomam o valor 1.0 e, conseqüentemente, o *logit* é zero. Pode-se verificar que *logits* negativos representam probabilidades abaixo de 0.5 e *logits* positivos representam probabilidades acima de 0.5.

Assim,

$$\text{logit} = \ln \left[\frac{P_i(x)}{Q_i(x)} \right] = B_0 + B_1X \quad (2.8)$$

e resolvendo em ordem de $P_i(X)$ tem-se

$$P_i(X) = \frac{e^{B_0+B_1X}}{1 + e^{B_0+B_1X}} \quad \text{ou} \quad P_i(X) = \frac{1}{1 + e^{-(B_0+B_1X)}}. \quad (2.9)$$

Nestas condições, pode-se afirmar que o modelo de regressão logística é um modelo linear assumindo o *logit* da probabilidade $P_i(X)$ e não a probabilidade em si.

No contexto da TRI, os modelos logísticos diferem um do outro pelo número de parâmetros que usam para descrever o item, nomeadamente: dificuldade do item, discriminação do item e a probabilidade de acerto casual (resposta correta dada por indivíduos com baixa habilidade). Tratam-se de modelos desenvolvidos para descrever a relação entre as características dos itens e o traço do respondente.

2.3.1 Modelo logístico de 1 parâmetro (1PL)

Este modelo considera apenas o parâmetro dificuldade do item, assumindo que todos os itens têm o mesmo parâmetro de discriminação igual a um. Este modelo é mais conhecido como modelo de Rasch (Baker, 1992) e é dado por:

$$P(Y_{ij} = 1|\theta_j) = \frac{e^{(\theta_j - b_i)}}{1 + e^{(\theta_j - b_i)}} = \frac{1}{1 + e^{-(\theta_j - b_i)}}, \quad (2.10)$$

onde $i = 1, 2, \dots, n$ e $j = 1, 2, \dots, J$, sendo:

$P(Y_{ij} = 1|\theta_j)$ - a probabilidade do indivíduo j com habilidade θ_j responder corretamente ao item i ;

Y_{ij} - uma variável dicotômica que assume o valor 1 quando o indivíduo j responde corretamente ao item i ou, 0 quando o indivíduo j não responde corretamente ao item i ;

θ_j - a habilidade do indivíduo j ;

b_i - o parâmetro da dificuldade do item i medido na mesma escala de medida da habilidade;

e - o número de Neper (base de logaritmo natural).

O parâmetro de dificuldade b , é um parâmetro de posição ou localização pois indica a posição na CCI em relação à escala da habilidade. De acordo com Andrade et al (2000) este parâmetro representa o nível da habilidade necessária para que a probabilidade de uma resposta

correta seja de $\frac{1+c}{2}$, ou simplesmente 50% neste caso pois $c = 0$, e, portanto, quanto maior o valor de b mais difícil é o item, e vice-versa.

A diferença $\theta_j - b_i$, na expressão (2.10) sugere que a probabilidade de responder corretamente é uma função da distância entre a habilidade do indivíduo e a dificuldade do item. Quando $\theta_j = b_i$, o indivíduo tem chance de 50% de responder corretamente ou não ao item. Quando $\theta_j > b_i$, o indivíduo tem maior probabilidade de responder corretamente ao item e, caso contrário, a probabilidade de acertar é menor. Quer a estimativa da habilidade θ , quer a estimativa do parâmetro de dificuldade b podem variar conforme o intervalo $] - \infty; \infty[$. É óbvio que para diferentes valores de θ e b se tenha, em algum momento, a mesma distância. Portanto, há necessidade de se definir a escala de medida para θ . Assumindo θ com distribuição normal de média zero e desvio-padrão 1, isto é, $\theta \sim N(0, 1)$, coloca-se o parâmetro dificuldade do item na mesma escala de distribuição normal. Nestas condições, os itens com valores negativos do parâmetro de dificuldade são tidos como relativamente fáceis ao passo que os itens com valores positivos do parâmetro de dificuldades são tidos como relativamente difíceis. A partir desta escala, os valores do parâmetro b variam entre -3 e $+3$ com uma elevada probabilidade (99,73%).

2.3.2 Modelo logístico de 2 parâmetros (2PL)

Neste modelo, os itens são descritos por meio de dois parâmetros: a dificuldade do item (visto na Subsecção 2.3.1) e a discriminação do item. Tem-se:

$$P(Y_{ij} = 1|\theta_j) = \frac{e^{a_i(\theta_j - b_i)}}{1 + e^{a_i(\theta_j - b_i)}} = \frac{1}{1 + e^{-a_i(\theta_j - b_i)}} \quad (2.11)$$

onde $i = 1, 2, \dots, n$ e $j = 1, 2, \dots, J$, sendo:

a_i - o parâmetro de discriminação do item, com valor proporcional à inclinação da CCI no ponto b_i .

O parâmetro de discriminação do item pode assumir tanto valores negativos bem como valores positivos porém, valores negativos para parâmetro de discriminação do item indicariam

que a probabilidade de um indivíduo responder corretamente ao item diminui com o aumento da sua habilidade, contrariando aquilo que é a CCI. Assim sendo, é aconselhável que a discriminação seja positiva, isto é, $a \in [0; +\infty[$ e, no caso de a habilidade assumir distribuição normal, $a \in [0; 3[$ com uma elevada probabilidade. Os itens que apresentam valores altos de discriminação têm as respectivas curvas características mais íngrimes e discriminam melhor os indivíduos pois fornecem mais informações sobre a habilidade.

A expressão $a_i(\theta_j - b_i)$ na equação (2.11) designa-se *logit*: logaritmo das odds (chance de responder corretamente ao item). Se

$$P(Y_{ij} = 1|\theta_j) = \frac{1}{1 + e^{-a_i(\theta_j - b_i)}} \quad \text{e} \quad 1 - P(Y_{ij} = 1|\theta_j) = \frac{1}{1 + e^{a_i(\theta_j - b_i)}} \quad (2.12)$$

então

$$\text{logit} = \ln \left[\frac{P(Y_{ij} = 1|\theta_j)}{1 - P(Y_{ij} = 1|\theta_j)} \right] = \ln e^{a_i(\theta_j - b_i)} = a_i(\theta_j - b_i). \quad (2.13)$$

2.3.3 Modelo logístico de 3 parâmetros (3PL)

Este é o modelo que integra todos os parâmetros dos itens: de dificuldade, de discriminação e de acerto casual. A forma matemática deste modelo é

$$P(Y_{ij} = 1|\theta_j) = c_i + (1 - c_i) \frac{1}{1 + e^{-a_i(\theta_j - b_i)}} \quad (2.14)$$

onde $i = 1, 2, \dots, n$ e $j = 1, 2, \dots, J$, sendo:

c_i - o parâmetro do item que indica a probabilidade de examinando com baixa habilidade responder corretamente o item.

O valor do parâmetro c varia entre 0 e 1 e corresponde ao ponto de intersecção entre a CCI (através da assíntota inferior) e o eixo das ordenadas. Itens com valores de c muito próximos de 0 são tidos como os melhores.

Por questão de simplificação, considere-se:

$$P_{ij} = P(Y_{ij} = 1|\theta_j) \quad (2.15)$$

$$Q_{ij} = 1 - P(Y_{ij} = 1|\theta_j) \quad (2.16)$$

2.3.4 Função de informação do item e função de informação do teste

Juntamente com a CCI, medidas como função de informação do item e função da informação do teste são utilizadas para, respetivamente, analisar a informação contida no item e no teste em relação ao traço latente em estudo (Reise et al, 2005). A função de informação do item, obtida a partir da informação de Fisher (I_F) - $I_F = E \left[\left(\frac{\partial \log f(Y_{ij}|\theta_j)}{\partial \theta_j} \right)^2 \right]$, é dada por

(i) Modelo 1PL:

$$I_i(\theta) = P_{ij} Q_{ij} \quad (2.17)$$

(ii) Modelo 2PL:

$$I_i(\theta) = a_i^2 P_{ij} Q_{ij} \quad (2.18)$$

(iii) Modelo 3PL:

$$I_i(\theta) = \frac{a_i^2 Q_{ij}}{P_{ij}} \left[\frac{P_{ij} - c_i}{1 - c_i} \right]^2 \quad (2.19)$$

onde $I_i(\theta)$ é a informação fornecida pelo item i no nível de habilidade θ .

A quantidade de informação fornecida por um item depende, fundamentalmente, dos valores dos respetivos parâmetros. Em geral, a informação é maior (i) quando b_i se aproxima de θ , (ii) quanto maior for a_i e (iii) quanto mais próximo c_i estiver de 0 (Andrade et al, 2000).

A soma das quantidades de informação fornecida por cada item que compõe o teste resulta na quantidade de informação fornecida pelo teste, isto é,

$$I(\theta) = \sum_{i=1}^n I_i(\theta). \quad (2.20)$$

2.4 Estimação dos Parâmetros

Nas secções anteriores vimos que a probabilidade de um indivíduo responder corretamente a um item depende dos parâmetros do item e da habilidade do indivíduo. Vimos ainda que esses parâmetros nem sempre são conhecidos, havendo necessidade de estimá-los. Aliás, esse é o grande constrangimento da TRI. O que se conhece são as respostas dos indivíduos aos itens. Assim, três casos podem se verificar aquando da aplicação dos modelos da TRI:

1. Estimar os parâmetros dos itens, sendo conhecidas as habilidades dos indivíduos;
2. Estimar as habilidades dos indivíduos, sendo conhecidos os parâmetros dos itens;
3. Estimar, simultaneamente, os parâmetros dos itens e as habilidades dos indivíduos.

Para qualquer um dos casos arrolados acima, a estimação dos parâmetros (dos itens e das habilidades) pode ser feita pelo método da máxima verosimilhança conjunta ou pelo método da máxima verosimilhança marginal para estimação de parâmetros em duas etapas sugerida por Bock & Lieberman (1970). Bock & Aitkin (1981) propuseram uma modificação ao método da máxima verosimilhança marginal permitindo a estimação dos parâmetros aplicando o algoritmo EM de Dempster (1977). Os métodos bayesianos são também utilizados para o mesmo fim (Baker, 1992; Baker, 2000) contudo, não serão discutidos neste trabalho.

Seja Y_{ij} uma variável aleatória associada à resposta binária dada pelo indivíduo j ao item i , com

$$Y_{ij} = \begin{cases} 1, & \text{resposta correta} \\ 0, & \text{resposta errada} \end{cases} \quad \text{onde } i = 1, \dots, n \text{ e } j = 1, \dots, J.$$

Sejam θ_j a habilidade do indivíduo j , η_i o conjunto dos parâmetros dos itens (a_i , b_i e c_i) e \mathbf{Y} o conjunto dos resultados possíveis da variável Y_{ij} . Para diferentes conjuntos de respostas possíveis \mathbf{Y} , a função a seguir pode ser usada para descrever as probabilidades com as quais Y_{ij} assume esses valores

$$f(y_{ij}|\theta_j, \eta_i) = P(Y_{ij} = y_{ij}|\theta_j, \eta_i), \quad y_{ij} \in \mathbf{Y}. \quad (2.21)$$

Se para um indivíduo j com habilidade θ_j são observadas respostas específicas $\mathbf{Y}_j = (y_{1j}, y_{2j}, \dots, y_{ij}, \dots, y_{nj})$ a probabilidade conjunta (tendo em conta a independência local) é dada pela função de verosimilhança

$$L(\theta_j, \eta | \mathbf{Y}_j) = \prod_{i=1}^n P(y_{ij} | \theta_j, \eta_i). \quad (2.22)$$

Se se tratar de todos os indivíduos que respondem ao teste, a função acima pode ser generalizada na forma

$$L(\theta, \eta | \mathbf{Y}) = \prod_{j=1}^J \prod_{i=1}^n P(y_{ij} | \theta_j, \eta_i) \quad \text{onde} \quad \theta = (\theta_1, \theta_2, \dots, \theta_J). \quad (2.23)$$

Uma vez que Y_{ij} segue uma distribuição de Bernoulli, a expressão (2.23) pode ser escrita na forma

$$L(\theta, \eta | \mathbf{Y}) = \prod_{j=1}^J \prod_{i=1}^n P(Y_{ij} = 1 | \theta_j)^{y_{ij}} [1 - P(Y_{ij} = 1 | \theta_j)]^{1-y_{ij}}. \quad (2.24)$$

Para cada resposta observada $Y_{ij} = y_{ij}$, e tendo em conta as expressões (2.15) e (2.16) tem-se a função de verosimilhança

$$L(\theta, \eta | \mathbf{y}) = \prod_{j=1}^J \prod_{i=1}^n P_{ij}^{y_{ij}} Q_{ij}^{1-y_{ij}}. \quad (2.25)$$

Os estimadores da máxima verosimilhança são os valores dos parâmetros que maximizam a função de verosimilhança. Portanto, deve-se determinar os zeros das derivadas parciais de primeira ordem. Por forma a evitar cálculos complexos com muitos termos de produtos, a sugestão é logaritmizar a função da verosimilhança. Seja $l(\theta, \eta | \mathbf{y}) = \log L(\theta, \eta | \mathbf{y})$. Assim,

$$l(\theta, \eta | \mathbf{y}) = \sum_{j=1}^J \sum_{i=1}^n [y_{ij} \ln P_{ij} + (1 - y_{ij}) \ln Q_{ij}]. \quad (2.26)$$

2.4.1 Estimação dos parâmetros dos itens sendo conhecidas as habilidades

Uma vez conhecidas as habilidades dos indivíduos que respondem ao teste e pretendendo-se estimar os parâmetros dos itens, a expressão (2.26) passa a depender de η :

$$l(\eta|\mathbf{y}) = \sum_{i=1}^J \sum_{j=1}^n y_{ij} \ln P_{ij} + (1 - y_{ij}) \ln Q_{ij} \quad (2.27)$$

A seguir são calculados as raízes da equação

$$\frac{\partial l(\eta|\mathbf{y})}{\partial \eta_i} = 0 \quad \text{onde} \quad \eta_i = (a_i, b_i, c_i), \quad (2.28)$$

$$\frac{\partial l(\eta|\mathbf{y})}{\partial \eta_i} = \sum_{j=1}^J \left[y_{ij} \frac{\partial \ln P_{ij}}{\partial \eta} + (1 - y_{ij}) \frac{\partial \ln Q_{ij}}{\partial \eta} \right] \quad (2.29)$$

$$= \sum_{j=1}^J \left[\frac{y_{ij}}{P_{ij}} - \frac{(1 - y_{ij})}{Q_{ij}} \right] \frac{\partial P_{ij}}{\partial \eta} \quad (2.30)$$

$$= \sum_{j=1}^J \left[\frac{y_{ij} - P_{ij}}{P_{ij} Q_{ij}} \right] \frac{\partial P_{ij}}{\partial \eta} \quad (2.31)$$

No caso geral dos modelos unidimensionais da TRI para itens dicotômicos,

$$P_{ij} = c_i + (1 - c_i) \frac{1}{1 + e^{-a_i(\theta_j - b_i)}} \quad \text{onde} \quad i = 1, 2, \dots, n; \quad j = 1, 2, \dots, J. \quad (2.32)$$

Aplicando algumas transformações algébricas em (2.32) tem-se:

$$Q_{ij} = 1 - P_{ij} = \frac{(1 - c_i)e^{-a_i(\theta_j - b_i)}}{1 + e^{-a_i(\theta_j - b_i)}} \quad (2.33)$$

$$P_{ij}^* = \frac{P_{ij} - c_i}{1 - c_i} = \frac{1}{1 + e^{-a_i(\theta_j - b_i)}} \quad (2.34)$$

$$Q_{ij}^* = \frac{Q_{ij}}{1 - c_i} = \frac{e^{-a_i(\theta_j - b_i)}}{1 + e^{-a_i(\theta_j - b_i)}} \quad (2.35)$$

As derivadas parciais da expressão (2.32) em ordem a a_i , b_i e c_i e levando em consideração as expressões (2.33) a (2.35), são indicadas a seguir:

$$\frac{\partial P_{ij}}{\partial a_i} = (\theta_j - b_i)(1 - c_i) P_{ij}^* Q_{ij}^* \quad (2.36)$$

$$\frac{\partial P_{ij}}{\partial b_i} = -a_i(1 - c_i) P_{ij}^* Q_{ij}^* \quad (2.37)$$

$$\frac{\partial P_{ij}}{\partial c_i} = Q_{ij}^* \quad (2.38)$$

Substituindo os valores das expressões (2.36) a (2.38) em (2.31) obtêm-se as equações que permitem encontrar os valores dos parâmetros a_i , b_i e c_i :

$$a_i : \quad \frac{\partial l(\eta|\mathbf{y})}{\partial a_i} = (1 - c_i) \sum_{j=1}^J (\theta_j - b_i)(y_{ij} - P_{ij})W_{ij} = 0 \quad (2.39)$$

$$b_i : \quad \frac{\partial l(\eta|\mathbf{y})}{\partial b_i} = -a_i(1 - c_i) \sum_{j=1}^J (y_{ij} - P_{ij})W_{ij} = 0 \quad (2.40)$$

$$c_i : \quad \frac{\partial l(\eta|\mathbf{y})}{\partial c_i} = \sum_{j=1}^J (y_{ij} - P_{ij}) \frac{W_{ij}}{P_{ij}^*} = 0 \quad (2.41)$$

sabendo que $W_{ij} = \frac{P_{ij}^* Q_{ij}^*}{P_{ij} Q_{ij}}$.

O sistema de equações da verosimilhança composto pelas equações (2.39), (2.40) e (2.41) para além de não ser linear, as suas equações não resultam em soluções explícitas de modo que é necessário recorrer a um procedimento iterativo como o caso do método de Newton-Raphson, por exemplo.

Considerando $[\hat{\eta}_i]_t$ uma estimativa de η_i na iteração t , o método de Newton-Raphson, com fórmula iteradora dada por:

$$[\hat{\eta}_i]_{t+1} = [\hat{\eta}_i]_t - [H(\hat{\eta}_i)]_t^{-1} \cdot [h(\hat{\eta}_i)]_t \quad (2.42)$$

requer os resultados para $H(\eta_i)$ e para $h(\eta_i)$. Os resultados que se seguem foram desenvolvidos por Andrade et al(2000):

$$\begin{aligned} H(\eta_i) &\equiv \frac{\partial^2 l(\eta)}{\partial \eta_i \partial \eta_i'} \\ &= \sum_{j=1}^J \left\{ \left(\frac{y_{ij} - P_{ij}}{P_{ij} Q_{ij}} \right) (P_{ij}^* Q_{ij}^*) H_{ij} - \left(\frac{y_{ij} - P_{ij}}{P_{ij} Q_{ij}} \right)^2 (P_{ij}^* Q_{ij}^*)^2 h_{ij} h_{ij}' \right\} \\ &= \sum_{j=1}^J (y_{ij} - P_{ij}) W_{ij} \{ H_{ij} - (y_{ij} - P_{ij}) W_{ij} h_{ij} h_{ij}' \} \end{aligned} \quad (2.43)$$

$$\begin{aligned}
h(\eta_i) &\equiv \frac{\partial l(\eta)}{\partial \eta_i} \\
&= \sum_{j=1}^J \left\{ (y_{ij} - P_{ij}) \frac{W_{ij}}{P_{ij}^* Q_{ij}^*} \right\} (P_{ij}^* Q_{ij}^*) h_{ij} \\
&= \sum_{j=1}^J (y_{ij} - P_{ij}) W_{ij} h_{ij}
\end{aligned} \tag{2.44}$$

$$\begin{aligned}
H_{ij} &= (P_{ij}^* Q_{ij}^*)^{-1} \left(\frac{\partial^2 P_{ij}}{\partial \eta_i \partial \eta_i'} \right) \\
&= \begin{pmatrix} (1 - c_i)(\theta_j - b_i)(1 - 2P_{ij}^*) & \cdot & \cdot \\ -(1 - c_i)\{1 + a_i(\theta_j - b_i)(1 - 2P_{ij}^*)\} & a_i^2(\theta_j - b_i)(1 - 2P_{ij}^*) & \cdot \\ -(\theta_j - b_i) & a_i & 0 \end{pmatrix}
\end{aligned} \tag{2.45}$$

e

$$h_{ij} = (P_{ij}^* Q_{ij}^*)^{-1} \left(\frac{\partial P_{ij}}{\partial \eta_i} \right) = \begin{pmatrix} (1 - c_i)(\theta_j - b_i) \\ -a_i(1 - c_i) \\ \frac{1}{P_{ij}^*} \end{pmatrix} \tag{2.46}$$

2.4.2 Estimação das habilidades sendo conhecidos os parâmetros dos itens

Neste caso, a função da verossimilhança (2.26) toma a forma

$$l(\theta|\mathbf{y}) = \sum_{i=1}^J \sum_{j=1}^n y_{ij} \ln P_{ij} + (1 - y_{ij}) \ln Q_{ij}. \tag{2.47}$$

Aplicando o raciocínio análogo para a obtenção dos zeros da derivada parcial de primeira ordem de (2.32) tem-se:

$$\frac{\partial l(\theta|\mathbf{y})}{\partial \theta_j} = \sum_{i=1}^n \left[y_{ij} \frac{\partial \ln P_{ij}}{\partial \theta_j} + (1 - y_{ij}) \frac{\partial \ln Q_{ij}}{\partial \theta_j} \right] \quad (2.48)$$

$$= \sum_{i=1}^n \left[\frac{y_{ij}}{P_{ij}} - \frac{(1 - y_{ij})}{Q_{ij}} \right] \frac{\partial P_{ij}}{\partial \theta_j} \quad (2.49)$$

$$= \sum_{i=1}^n \left[(y_{ij} - P_{ij}) \frac{1}{P_{ij} Q_{ij}} \right] \frac{\partial P_{ij}}{\partial \theta_j} \quad (2.50)$$

$$= \sum_{i=1}^n \left[(y_{ij} - P_{ij}) \frac{W_{ij}}{P_{ij}^* Q_{ij}^*} \right] \frac{\partial P_{ij}}{\partial \theta_j} \quad (2.51)$$

Mas como

$$\frac{\partial P_{ij}}{\partial \theta_j} = a_i(1 - c_i) P_{ij}^* Q_{ij}^* \quad (2.52)$$

então,

$$\begin{aligned} \frac{\partial l(\theta|\mathbf{y})}{\partial \theta_j} &= \sum_{i=1}^n a_i(1 - c_i)(y_{ij} - P_{ij})W_{ij} \quad \text{e, portanto} \\ \theta_j : \sum_{i=1}^n a_i(1 - c_i)(y_{ij} - P_{ij})W_{ij} &= 0. \end{aligned} \quad (2.53)$$

Como é óbvio, a equação (2.53) requer aplicação de procedimentos iterativos para a obtenção da solução. A seguir mostra-se, mais uma vez, como o método de Newton-Raphson pode ser útil para obter a estimativa do parâmetro da habilidade. A expressão

$$\left[\hat{\theta}_j \right]_{t+1} = \left[\hat{\theta}_j \right]_t - \left[\frac{\partial^2 l(\theta)}{\partial \theta_j^2} \right]_t^{-1} \cdot \left[\frac{\partial l(\theta)}{\partial \theta_j} \right]_t \quad (2.54)$$

quando aplicada, permite obter um estimador para a habilidade $\hat{\theta}_j$ do examinando. O cálculo da segunda derivada que consta da expressão (2.54) que, na verdade, corresponde à matriz hessiana em função do θ_j , leva ao seguinte resultado desenvolvido por Andrade et al (2000):

$$\begin{aligned} H(\theta_j) &\equiv \frac{\partial^2 l(\theta)}{\partial \theta_j^2} \\ \frac{\partial^2 l(\theta)}{\partial \theta_j^2} &= \sum_{i=1}^n (y_{ij} - P_{ij}) W_{ij} \{ H_{ij} - (y_{ij} - P_{ij}) W_{ij} h_{ij}^2 \} \\ \frac{\partial^2 l(\theta)}{\partial \theta_j^2} &= \sum_{i=1}^n (y_{ij} - P_{ij}) W_{ij} \{ H_{ij} - (y_{ij} - P_{ij}) W_{ij} h_{ij}^2 \} \end{aligned} \quad (2.55)$$

onde

$$H_{ij} = (P_{ij}^* Q_{ij}^*)^{-1} \left(\frac{\partial^2 P_{ij}}{\partial \theta_j^2} \right) = a_i^2 (1 - c_i) (1 - 2P_{ij}^*) \quad (2.56)$$

$$h_{ij} = (P_{ij}^* Q_{ij}^*)^{-1} \left(\frac{\partial P_{ij}}{\partial \theta_j} \right) = a_i (1 - c_i). \quad (2.57)$$

De acordo com a expressão (2.50), $\frac{\partial l(\theta)}{\partial \theta_j} = \sum_{i=1}^n \left[(y_{ij} - P_i) \frac{1}{P_i Q_i} \right] \frac{\partial P_i}{\partial \theta_j}$ mas, tendo em conta o resultado (2.52), tem-se

$$\frac{\partial l(\theta)}{\partial \theta_j} = \sum_{i=1}^n (y_{ij} - P_{ij}) W_{ij} h_{ij}. \quad (2.58)$$

Substituindo as expressões (2.55) e (2.58) na expressão (2.54) obtém-se

$$\left[\hat{\theta}_j \right]_{t+1} = \left[\hat{\theta}_j \right]_t + \left[\frac{\sum_{i=1}^n (y_{ij} - P_{ij}) W_{ij} h_{ij}}{\sum_{i=1}^n (y_{ij} - P_{ij}) W_{ij} \{ H_{ij} - (y_{ij} - P_{ij}) W_{ij} h_{ij}^2 \}} \right]_t. \quad (2.59)$$

A expressão (2.59) é resolvida iterativamente para o valor de $\hat{\theta}_j$, para cada examinando.

2.4.3 Estimação dos parâmetros dos itens e das habilidades

Este é o caso mais comum na TRI. Os parâmetros dos modelos propostos são estimados através das técnicas de estimação da máxima verosimilhança conjunta (EMVC) ou estimação da máxima verosimilhança marginal (EMVM). No caso de EMVC, apesar das habilidades dos examinandos não serem conhecidas, estas devem ser estimadas juntamente com os parâmetros dos itens. O que se faz é (i) supor serem conhecidas as habilidades dos examinandos e a partir daí estimar os parâmetros dos itens e (ii) com os parâmetros dos itens estimados em (i), estimar em seguida as habilidades dos examinandos. O processo é repetido tantas quantas vezes forem necessárias até se atingir a convergência, aplicando um método iterativo como o de Newton-Raphson, abordado nas Subsecções 3.4.1 e 3.4.2. Contudo, a implementação deste método tem limitações computacionais devido à dimensão da matriz hessiana necessária para efetuar os cálculos para além de que os parâmetros dos itens tidos como estruturais (dimensão do teste) e as habilidades dos examinandos tidas como incidentais (dimensão da amostra), quando estimados em simultâneo as estimativas da máxima verosimilhança não

são consistentes agravando-se com o aumento da amostra (Baker, 1992). Assim, por forma a superar essas limitações, é sugerida a técnica de EMVM tida como computacionalmente factível.

A técnica de estimação de parâmetros via EMVM foi proposta por Bock & Lieberman (1970) e mais tarde desenvolvida por Bock & Aitkin (1981). Comparativamente à técnica de EMVC, a EMVM é considerada mais vantajosa por apresentar resultados teoricamente aceites e por ser computacionalmente factível. A EMVM consiste em assumir a existência de uma distribuição de probabilidade associada às habilidades e considerar que os examinandos representam uma amostra dessa distribuição. Essa distribuição tem como função densidade $g(\theta|\mathcal{T})$, onde \mathcal{T} é o vetor de parâmetros associado à distribuição g da habilidade. Embora θ possa apresentar várias distribuições, o caso mais comum é quando assume uma distribuição normal. Importa referir que a EMVM recorre ao Teorema de Bayes para efeitos matemático não significando necessariamente se tratar de um método Bayesiano.

2.4.4 Abordagem de Bock & Lieberman

Com os pressupostos da técnica de EMVM, θ passa a ser aleatório e, portanto, a probabilidade marginal de um examinando j apresentar um determinado padrão de respostas $y_{.j}$ será dada por:

$$P(Y_{.j} = y_{.j}|\eta) = \int P(y_{.j}|\theta, \eta) g(\theta|\mathcal{T}) d\theta. \quad (2.60)$$

Aplicando o Teorema de Bayes para definir a distribuição de θ_j dado o vetor $y_{.j}$, tem-se:

$$P(\theta_j|y_{.j}, \eta, \mathcal{T}) = \frac{P(y_{.j}|\theta_j, \eta) g(\theta|\mathcal{T})}{\int P(y_{.j}|\theta_j, \eta) g(\theta|\mathcal{T}) d\theta}. \quad (2.61)$$

Assumindo a independência local, a probabilidade associada ao vetor das respostas (primeiro termo da expressão (2.61) no numerador) é dada por

$$P(y_{.j}|\theta_j, \eta) = \prod_{i=1}^n P_{ij}^{y_{ij}} Q_{ij}^{1-y_{ij}} \quad (2.62)$$

que na verdade, é a função de verosimilhança.

No contexto de Bock & Lieberman (1970), a função de verosimilhança marginal é dada por

$$L = \prod_{j=1}^J P(y_{.j}|\theta_j, \eta). \quad (2.63)$$

Por conseguinte, a respectiva função logarítmica será

$$l = \log(L) = \sum_{j=1}^J \ln P(y_{.j}|\theta_j, \eta) \quad (2.64)$$

e, naturalmente, as estimativas que maximizam l são encontradas quando

$$\frac{\partial l}{\partial \eta} = 0, \quad \text{isto é,} \quad \frac{\partial l}{\partial a_i} = 0; \quad \frac{\partial l}{\partial b_i} = 0; \quad \frac{\partial l}{\partial c_i} = 0. \quad (2.65)$$

Resolvendo as equações (2.65) (ver os desenvolvimentos em Baker (1992)) são obtidas as equações da verosimilhança marginal, respetivamente, para a_i , b_i e c_i :

$$\frac{\partial l}{\partial a_i} = (1 - c_i) \sum_{j=1}^J \int [(\theta_j - b_i)(y_{ij} - P_{ij})W_{ij}] P(\theta_j|y_{.j}, \eta, \tau) d\theta = 0 \quad (2.66)$$

$$\frac{\partial l}{\partial b_i} = -a_i(1 - c_i) \sum_{j=1}^J \int [(y_{ij} - P_{ij})W_{ij}] P(\theta_j|y_{.j}, \eta, \tau) d\theta = 0 \quad (2.67)$$

$$\frac{\partial l}{\partial c_i} = \frac{1}{1 - c_i} \sum_{j=1}^J \int \left[\frac{y_{ij} - P_{ij}}{P_{ij}} \right] P(\theta_j|y_{.j}, \eta, \tau) d\theta = 0 \quad (2.68)$$

A partir das equações de verosimilhança marginal (2.66) a (2.68) são estimados os parâmetros dos itens que seguidamente são utilizados para atualizar a informação da distribuição de θ e portanto os parâmetros dos itens são, novamente, re-estimados. O procedimento (com recurso a um método iterativo) é repetido até que os valores estimados estabilizem. Uma vez estabilizadas as estimativas dos parâmetros dos itens e determinada a distribuição de θ , então a pontuação θ para cada examinando pode ser estimada. O método iterativo a ser utilizado deve permitir encontrar uma aproximação numérica que envolvem integrais e para tal sugere-se o método de Hermite-Gauss ou, simplesmente, o método da quadratura.

Na perspectiva da quadratura, se $g(\theta|\mathcal{T})$ for uma distribuição contínua com momentos finitos, ela pode ser aproximada para qualquer grau de precisão, por uma outra distribuição que assume um número finito de pontos (um histograma, por exemplo). Assim, o problema de obter o integral de uma distribuição contínua (área sob a curva) é substituída pela obtenção da soma das áreas de um número finito de q retângulos que se aproxima à área sob a curva. O ponto médio de cada retângulo, na escala da habilidade, X_k ($k = 1, 2, \dots, q$) é chamado de “nó” ou ponto de quadratura. Cada nó tem um peso associado $A(X_k)$ que leva em conta a altura da densidade $g(\theta|\mathcal{T})$ na vizinhança de X_k e a largura dos retângulos. Os valores de X_k e $A(X_k)$, para aproximação da curva gaussiana mas não com distribuição normal univariada, já se encontram tabelados e fornecidos por Stroud & Secrest (1996). Para se obter uma distribuição aproximada a g , os valores tabelados de X_k são multiplicados por $\sqrt{2}$ e os pesos $A(X_k)$, divididos por $\sqrt{\pi}$. No caso de $g(\theta|\mathcal{T})$, não é necessário que tenha distribuição normal pois, em geral, pode ser definida empiricamente.

Substituindo o nível de habilidades conhecidos θ_j pelos q valores de X_k , a relação de Bock & Lieberman (2.61), escrita na forma da quadratura, fica

$$P(X_k|y_{.j}, \eta, \mathcal{T}) = \frac{P(y_{.j}|X_k, \eta) A(X_k)}{\sum_{k=1}^q P(y_{.j}|X_k, \eta) A(X_k)}. \quad (2.69)$$

Utilizando a aproximação da quadratura 2.69 para o integral nas equações (2.66) a (2.68) tem-se

$$a_i : \quad (1 - c_i) \sum_{k=1}^q \sum_{j=1}^J [y_{ij} - P_{ik}(X_k)] (X_k - b_i) W_{ik} P(X_k|y_{.j}, \eta, \tau) = 0 \quad (2.70)$$

$$b_i : \quad -a_i(1 - c_i) \sum_{k=1}^q \sum_{j=1}^J (y_{ij} - P_{ik}(X_k)) W_{ij} P(X_k|y_{.j}, \eta, \tau) = 0 \quad (2.71)$$

$$c_i : \quad \frac{1}{1 - c_i} \sum_{k=1}^q \sum_{j=1}^J \left[\frac{y_{ij} - P_{ik}(X_k)}{P_{ik}(X_k)} \right] P(X_k|y_{.j}, \eta, \tau) = 0 \quad (2.72)$$

onde

$$W_{ik} = \frac{P_{ik}^*(X_k) Q_{ik}^*(X_k)}{P_{ik}(X_k) Q_{ik}(X_k)} \quad (2.73)$$

e

$$P_{ik} = c_i + (1 - c_i) \frac{1}{1 + e^{-a_i(X_k - b_i)}}. \quad (2.74)$$

As equações (2.70), (2.71) e (2.72) não apresentam soluções explícitas e portanto, necessitam de aplicação do método de Newton-Raphson. Contudo, o método iterativo de Newton-Raphson não é muito bom computacionalmente por requerer a inversão de uma matriz $3n \times 3n$. Consequentemente, o processo de estimação é limitado a um número muito pequeno de itens. Ademais, as estimativas dos parâmetros dos itens não são independentes da amostra uma vez que o método requer que a distribuição da habilidade da amostra seja previamente conhecida. A proposta de Bock & Aitkin (1981) é considerada a solução para esse problema computacional.

2.4.5 Abordagem de Bock & Aitkin

Conforme referido na introdução da Secção 2.4.3, Bock & Aitkin (1981) reformularam as equações de verosimilhança de Bock & Lieberman tornando-as computacionalmente fatível e sob pressuposto de que a distribuição da população é conhecida ou presentemente estimada com as especificações corretas, produzem estimativas consistentes para os parâmetros dos itens. Bock & Aitkin (1981) assumem que os itens são independentes, as habilidades dos examinandos são independentes e os itens e as habilidades, também o são. Com estas condições, os parâmetros dos itens podem ser estimados um de cada vez com a habilidade do examinando a ser estimado por examinando.

Tomando como exemplo a equação de verosimilhança (2.70) para a_i , algumas alterações são recomendadas por forma a serem aplicados outros algoritmos que facilitem os cálculos. Assim, multiplicando pelo termo $P(X_k|y_{.j}, \eta, \mathcal{I})$ e distribuindo pelo somatório em função do índice j , tem-se

$$a_i : (1 - c_i) \sum_{k=1}^q (X_k - b_i) \left[\sum_{j=1}^J y_{ij} P(X_k|y_{.j}, \eta, \mathcal{I}) - P_{ik}(X_k) \sum_{j=1}^J P(X_k|y_{.j}, \eta, \mathcal{I}) \right] W_{ik} = 0 \quad (2.75)$$

Colocando a expressão (2.61) na sua forma de quadratura vem

$$P(X_k|y_{.j}, \eta, \mathcal{I}) = \frac{\prod_{i=1}^n P_{ik}^{y_{ij}} Q_{ik}^{1-y_{ij}} A(X_k)}{\sum_{k=1}^q \prod_{i=1}^n P_{ik}^{y_{ij}} Q_{ik}^{1-y_{ij}} A(X_k)} \quad (2.76)$$

que é a probabilidade *à posteriori* de um examinando ter o nível de habilidade X_k .

Fazendo

$$f_{ik} = \sum_{j=1}^J P(X_k|y_{.j}, \eta, \mathcal{I}) = \sum_{j=1}^J \frac{\prod_{i=1}^n P_{ik}^{y_{ij}} Q_{ik}^{1-y_{ij}} A(X_k)}{\sum_{k=1}^q \prod_{i=1}^n P_{ik}^{y_{ij}} Q_{ik}^{1-y_{ij}} A(X_k)} \quad (2.77)$$

e

$$r_{ik} = \sum_{j=1}^J y_{ij} P(X_k|y_{.j}, \eta, \mathcal{I}) = \sum_{j=1}^J \frac{\prod_{i=1}^n y_{ij} P_{ik}^{y_{ij}} Q_{ik}^{1-y_{ij}} A(X_k)}{\sum_{k=1}^q \prod_{i=1}^n P_{ik}^{y_{ij}} Q_{ik}^{1-y_{ij}} A(X_k)} \quad (2.78)$$

onde f_{ik} é interpretado como o número de examinandos de um total de J que se espera ter com o nível de habilidade X_k e r_{ik} é o número de examinandos com nível de habilidade X_k que se espera que respondam corretamente ao item i .

Seja, no contexto de Bock & Aitkin,

$$L(X_k) = \prod_{i=1}^n P_{ik}^{y_{ij}} Q_{ik}^{1-y_{ij}} \quad (2.79)$$

a quadratura da probabilidade condicional de $y_{.j}$, dados $\theta_j = X_k$ e os parâmetros dos itens. Então, a forma de quadratura para as expressões (2.76), (2.77) e (2.78) são, respetivamente,

$$P(X_k|y_{.j}, \eta, \mathcal{I}) = \frac{L(X_k) A(X_k)}{\sum_{k=1}^q L(X_k) A(X_k)} \quad (2.80)$$

$$f_{ik} = \sum_{j=1}^J \frac{L(X_k) A(X_k)}{\sum_{k=1}^q L(X_k) A(X_k)} \quad (2.81)$$

e

$$r_{ik} = \sum_{j=1}^J \frac{y_{ij} L(X_k) A(X_k)}{\sum_{k=1}^q L(X_k) A(X_k)}. \quad (2.82)$$

Nestas condições (tendo os valores de f_{ik} e de r_{ik}) a quadratura das equações de verosimilhança marginal são apresentadas, a seguir:

$$a_i : (1 - c_i) \sum_{k=1}^q (X_k - b_i) [r_{ik} - f_{ik} P_{ik}(X_k)] W_{ik} = 0 \quad (2.83)$$

$$b_i : -a_i(1 - c_i) \sum_{k=1}^q [r_{ik} - f_{ik} P_{ik}(X_k)] W_{ik} = 0 \quad (2.84)$$

$$c_i : \frac{1}{1 - c_i} \sum_{k=1}^q \left[\frac{r_{ik} - f_{ik} P_{ik}(X_k)}{P_{ik}(X_k)} \right] = 0. \quad (2.85)$$

As equações (2.83), (2.84) e (2.85) são equações de verosimilhança para uma análise probit para o item i na qual X_k é a variável independente. Diferentemente das equações encontradas na Secção 2.4.2, onde são conhecidos os valores de f_{ik} e r_{ik} , na solução de EMVM estes valores são desconhecidos e o número esperado de examinandos que respondem ao item i e o número esperado de respostas corretas são substituídos por essas quantidades desconhecidas em cada nó da quadratura (X_k). Estas quantidades são conhecidas na literatura da TRI como “dados artificiais” visto serem resultados das equações (2.77) e (2.78). Assim como apresentado na Secção 2.4.1, os parâmetros dos itens são estimados utilizando métodos iterativos tal como o de Newton-Raphson. No estágio da aplicação de EMVC para a estimação dos parâmetros dos itens, f_{ik} e r_{ik} são conhecidos. Porém, sob abordagem de EMVM os valores de f_{ik} e r_{ik} dependem dos valores das estimativas dos parâmetros dos itens. O que se sabe é que as equações (2.83), (2.84) e (2.85) não satisfazem essa dependência. Consequentemente, deve-se estabelecer um paradigma no qual tanto os “dados artificiais” quanto as estimativas dos parâmetros dos itens, possam ser obtidos iterativamente. É neste cenário em que se aplica, à solução de Bock & Aitkin, o algoritmo EM de Dempster (1977).

2.4.6 Algoritmo EM

O algoritmo EM é uma ferramenta computacional utilizada para o cálculo de estimativas de máxima verosimilhança de parâmetros de modelos probabilísticos de forma iterativa, na presença de variáveis aleatórias não observadas. É principalmente utilizado em problemas envolvendo dados incompletos ou valores em falta. Cada iteração deste processo é feito em

dois passos: o passo E (*Expectation*) e o passo M (*Maximization*). Recorde-se que o que se pretende na TRI é, a partir da habilidade dos examinandos (θ_j - variável não observada), estimar os parâmetros dos itens ($\eta_i = a_i, b_i, c_i$). Ademais, a inferência sobre θ_j é feita a partir das respostas observadas aos itens. Casella & Berger (2001) afirmam que o EM é um algoritmo projetado não só para encontrar os estimadores da máxima verosimilhança como também, seguramente, converge para os estimadores da máxima verosimilhança. A ideia base é a de substituir uma difícil maximização da verosimilhança por uma sequência de maximizações mais fáceis, cujo limite é a resposta para o problema original.

Seja Y o conjunto de dados observados incompletos e (Y, θ_j) o conjunto dos dados completos não observados e, ainda, $f(Y, \theta_j | \eta)$ a densidade conjunta dos dados completos. Dada a matriz provisória dos parâmetros dos itens na p -ésima iteração, η^{p+1} é calculado pela maximização da esperança do logaritmo da verosimilhança dos dados completos em relação a a, b, c , isto é, $E[\log f(Y, \theta_j | \eta) | Y, \eta^p]$. Note que η^{p+1} é o η obtido na p -ésima iteração do algoritmo. A seguir são apresentados os dois passos do algoritmo EM:

Passo-E: Calcular $E[\log f(Y, \theta_j | \eta) | Y, \eta^p]$;

Passo-M: Escolher η^{p+1} tal que a esperança é maximizada.

Os passos E e M são repetidos até que algum critério de convergência pré-estabelecido seja satisfeito.

Existem três formas do algoritmo EM, distinguidas pelas particulares restrições colocadas ao modelo probabilístico, isto é, a relação entre o modelo da TRI (probabilidade) e a distribuição da família exponencial: (i) quando a densidade $f(Y, \theta_j | \eta)$ é um membro regular da família exponencial (caso do modelo de Rasch), (ii) quando a densidade $f(Y, \theta_j | \eta)$ não é um membro regular da família exponencial mas um membro da família exponencial curvada e (iii) quando a densidade $f(Y, \theta_j | \eta)$ não tem nenhuma relação com a família exponencial (modelos 2PL e 3PL). Para a primeira forma, o procedimento torna-se relativamente simples pois a estatística para θ_j existe e o algoritmo reduz-se a tomar a esperança da suficiente

condicionada aos dados observados e os parâmetros provisórios estimados (Passo-E), substituindo essas esperanças condicionadas na maximização (Passo-M) e executar a habitual estimação da máxima verosimilhança. Contudo, o algoritmo EM pode ser aplicado, inclusive, por modelos que não sejam nenhum tipo de família exponencial.

No caso dos modelos logísticos 2PL e 3PL para sua aplicação, a distribuição de $f(Y, \theta_j | \eta)$ não é conhecida e, não sendo membro da família exponencial, a estatística suficiente da máxima verosimilhança não está disponível. Como substituto, os valores esperados do $\log f(Y, \theta_j | \eta)$, condicionados a algumas representações observadas de θ , são tomados e essas quantidades são tratadas como se fossem conhecidas (Passo-E). Esses valores esperados são utilizados para encontrar as estimativas dos parâmetros dos itens maximizando o logaritmo da função de verosimilhança (Passo-M) empregando métodos de máxima verosimilhança.

A seguir é descrito o algoritmo EM aplicado a TRI. Suponha que as habilidades estão restritas a um conjunto de q valores θ_k , com probabilidades $\pi_k, k = 1, \dots, q$. Seja f_{ki} o número de indivíduos com habilidade θ_k respondendo ao item i , $f_i = (f_{1i}, \dots, f_{ki}, \dots, f_{qi})'$ e $r_i = (r_{1i}, \dots, r_{ki}, \dots, r_{qi})'$. Se as habilidades dos J examinandos ($\sum_{k=1}^q f_{ki} = J$) constituírem uma amostra aleatória da distribuição acima, a probabilidade conjunta de que $f_{1i}, \dots, f_{ki}, \dots, f_{qi}$ estudantes terão níveis de habilidades $\theta_1, \dots, \theta_k, \dots, \theta_q$ é dada por uma distribuição multinomial

$$f(f_i | \pi) = \left[\frac{f!}{f_{1i}! \dots f_{ki}! \dots f_{qi}!} \right] \prod_{k=1}^q \pi_k^{f_{ki}}. \quad (2.86)$$

Dados f_{ki} e θ_k , a probabilidade do vetor de respostas, r_i , tem distribuição binomial de parâmetros f_{ki} e $P_{ki}(\theta_k)$, isto é,

$$\prod_{i=1}^n \binom{f_{ki}}{r_{ki}} P_i^{r_{ki}}(\theta_k) Q_i^{f_{ki}-r_{ki}}(\theta_k) \quad (2.87)$$

e a probabilidade conjunta dos vetores f e r é

$$\prod_{k=1}^q \prod_{i=1}^n \binom{f_{ki}}{r_{ki}} P_i^{r_{ki}}(\theta_k) Q_i^{f_{ki}-r_{ki}}(\theta_k) \left[\frac{f!}{f_{1i}! \dots f_{ki}! \dots f_{qi}!} \right] \prod_{k=1}^q \pi_k^{f_{ki}}. \quad (2.88)$$

A partir do critério de fatorização pode-se demonstrar que (f, r) é uma estatística suficiente para os dados completos (Y, θ) .

Ignorando os termos constantes, a função logaritmo da verosimilhança para os dados completos pode ser escrito na forma

$$\log(L) \propto \sum_{k=1}^q \sum_{i=1}^n \left[r_{ki} \ln P_i(\theta_k) + (f_{ki} - r_{ki}) \ln Q_i(\theta_k) + \sum_{k=1}^q f_{ki} \ln \pi_k \right] \quad (2.89)$$

Aqui (f, r) não é observado mas tomando a esperança de $\log(L)$ condicional em Y , dado η , tem-se:

$$E[\log(L)] = \sum_{k=1}^q \sum_{i=1}^n \left\{ E(r_{ki}|Y, \eta) \ln P_i(\theta_k) + E[(f_{ki} - r_{ki})|Y, \eta] \ln Q_i(\theta_k) + \sum_{k=1}^q E(f_{ki}|Y) \ln \pi_k \right\} \quad (2.90)$$

A última parcela de (2.90) pode ser ignorado uma vez que não depender de η . Maximizar (2.90) é equivalente a maximizar a expressão do Passe-E na apresentação do algoritmo EM e, para um dado $g(\theta|\mathcal{I})$ e um modelo TRI, proporciona as estimativas dos parâmetros dos itens que resolvem as equações da máxima verosimilhança marginal (2.83), (2.84) e (2.85).

Uma vez que se assume a independência dos itens, as segundas derivadas cruzadas de diferentes itens são zero no Passo-M e, portanto a maximização de $E[\log(L)]$ é realizada para cada item isoladamente. Efetivamente o que se faz é:

1. Passo-E:

- (a) Utilizar a quadratura $L(X_k) = \prod_{i=1}^n P_{ik}^{y_{ij}} Q_{ik}^{1-y_{ij}}$ e as estimativas provisórias para os parâmetros dos itens para calcular a verosimilhança do vector das pontuações de cada examinando em cada um dos q nós;
- (b) Utilizar a expressão $P(X_k|y_{\cdot j}, \eta, \mathcal{I}) = \frac{L(X_k) A(X_k)}{\sum_{k=1}^q L(X_k) A(X_k)}$ para obter a probabilidade *à posteriori* no ponto X_k com peso de quadratura $A(X_k)$ em cada um dos q nós;

- (c) Calcular o número esperado de examinandos que respondem ao item i e o número esperado de respostas corretas para esse item em cada um dos q nós pelas expressões $f_{ik} = \sum_{j=1}^J \frac{L(X_k) A(X_k)}{\sum_{k=1}^q L(X_k) A(X_k)}$ e $r_{ik} = \sum_{j=1}^J \frac{y_{ij} L(X_k) A(X_k)}{\sum_{k=1}^q L(X_k) A(X_k)}$, respetivamente.
2. Passo-M: com f_{ki} e r_{ki} obtidos no Passo-E, resolver as equações da máxima verosimilhança (2.83), (2.84) e (2.85) em relação a a_i , b_i e c_i usando o algoritmo de Newton-Raphson;
3. Se a função de máxima verosimilhança é inalterada em relação ao ciclo anterior, o processo de estimação do item converge e termina. Caso contrário, repetir 1 e 2.

Capítulo 3

Aplicação

Neste capítulo são apresentadas as análises realizadas a dados relativos ao desempenho dos alunos do 1^o ano de licenciatura em Biologia e os alunos do 2^o ano de licenciatura em Biologia e Geologia, na unidade curricular de Bioestatística, da responsabilidade do Departamento de Matemática da Universidade de Aveiro, no primeiro semestre civil do ano de 2016 (dados1), juntamente com os dados resultantes da apreciação e avaliação feita por alguns professores inquiridos, aos testes utilizados para avaliar os alunos (dados2).

Em relação aos dados recolhidos dos alunos (dados1), importa dizer que, durante o semestre mencionado acima, os alunos foram submetidos a cinco momentos de avaliação envolvendo os conteúdos programáticos da unidade curricular de Bioestatística (ver Anexo I). As questões (itens) que compõem as avaliações são, em cerca de 92%, do tipo escolha múltipla com apenas uma opção correta, dentre quatro opções propostas. Os momentos de avaliação 1 e 2 têm, respetivamente, 1 e 2 questões abertas. Estas questões foram dicotomizadas corrigindo-as com o critério de certo (percentagem de acerto igual ou acima de 60%) ou errado (percentagem de acerto abaixo de 60%). Finalmente os cinco momentos de avaliação foram compilados num único teste o qual foi utilizado para o presente estudo (ver Anexo II).

Relativamente aos dados recolhidos de professores (dados2), foi elaborado um questionário (ver o Apêndice B) através do qual 6 professores apreciaram e avaliaram o teste aplicado aos alunos. A avaliação solicitada foi delicada e minuciosa na medida em que os professores

tinham de analisar o teste item a item, o que pressupunha prévia resolução do teste, e só depois fariam as respectivas ilações. Por forma a evitar enviesamento dos resultados, os professores que participaram na elaboração e na aplicação do teste foram excluídos do inquérito.

Foi utilizado o *software* estatístico R (R Core Team, 2014) para efeitos de programação, devido à sua flexibilidade e acesso livre. A análise dos dados¹ foi feita sustentada na Teoria Clássica dos Testes (na análise descritiva) e na Teoria de Resposta ao Item (na seleção do melhor modelo e estimação dos respetivos parâmetros), e realizada com recurso ao pacote *ltm* do R (Rizopoulos, 2006). Recorreu-se, ainda dentro do R, o pacote *irr* (Gamer, 2015) para analisar o nível de concordância dos professores inquiridos quanto à validade e consistência do teste.

3.1 Descrição dos dados¹ via TCT

Na Secção 2.1 foi apresentada, de forma resumida e concisa, a TCT, os seus pressupostos e a sua aplicação na avaliação da qualidade de um teste. A partir da TCT foi possível obter, em relação ao teste, estatísticas como índice de dificuldades (proporção de acerto ao item), índice de discriminação (através do coeficiente de correlação ponto-bisserial) e o coeficiente de consistência interna (através do alpha de Cronbach). Os valores destas estatísticas por item podem ser vistos na Tabela (A.1) do Apêndice A. Observa-se que os índices de dificuldade variam entre 28.83% (item 2) e 95.50% (item 3). Obviamente que, com esses valores, o item 3 é considerado como o mais fácil e o item 2 o mais difícil. Relativamente à discriminação, com excepção do item 2 que tem valor negativo (-0.0147), todos os itens apresentam valores positivos variando entre 0.0357 (item 31) e 0.6586 (item 18). Embora o item 18 seja o mais discriminativo, os restantes, na sua maioria, têm boa discriminação. Valores negativos não são recomendados para o índice de discriminação pois mostra incoerência e inexistência de alinhamento entre o resultado do item e o resultado do teste. Nesses casos, o aconselhável é eliminar o item da análise. Em relação à consistência interna, os valores do alpha de Cronbach, eliminado cada um dos itens do teste, estão muito próximos do valor do alpha de Cronbach quando se consideram todos itens (0.7987), que é um valor aceitável. Essa

proximidade e semelhança dos valores mostram a adequação do teste. Eliminando o item 2 (o mais difícil e com discriminação negativa), a consistência interna aumenta para 0.8074. Outros itens que permitem aumentar a farsquia da consistência interna do teste para pelo menos 0.80, sem no entanto ultrapassar o caso que envolve o item 2, são os itens 1, 15, 21 e 31.

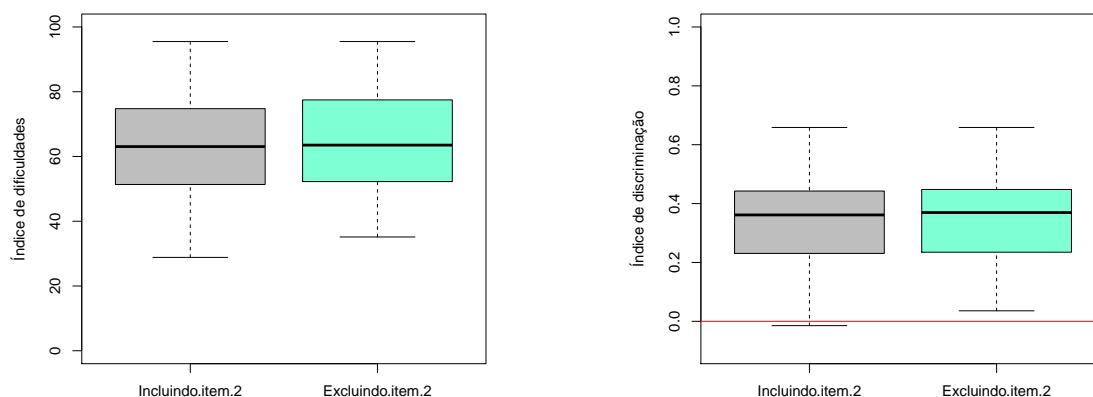
De um modo geral, os resultados mostram que o teste é composto por itens discriminativos e muito discriminativos, que existem itens de todos os níveis de dificuldade e com boa consistência interna. Portanto, há motivos para concordar que o teste produz os resultados esperados, isto é, mede as habilidades dos alunos na unidade curricular de Bioestatística.

Tendo em conta as constatações acima, é óbvio que o item 2 não é recomendável ao teste devendo ser revisto ou removido. A Tabela (3.1.1) apresenta os valores médios e globais para estas estatísticas envolvendo duas situações: (A) incluindo o item 2 e (B) excluindo o item 2. As colunas da tabela são, no contexto da TCT, ID o índice de dificuldade, Id o índice de discriminação, DP o desvio padrão e α o alpha de Cronbach. Apesar de ambas as situações apresentarem valores aceitáveis (ver as Tabelas (2.1.2) e (2.1.3)) verifica-se uma ligeira melhoria ao optar pela exclusão do item 2 não só nos valores do número de respostas corretas, ID, Id e α , como também no valores dos respetivos desvios padrão.

Situação	Alunos	Nº de Acertos (DP)	ID (DP)	Id (DP)	α
A	111	70,62 (18,96)	63,62 (17,08)	0,3441 (0,1563)	0,7987
B	111	71,69 (18,05)	64,59 (16,26)	0,3541 (0,1461)	0,8074

Tabela 3.1.1: Resumo descritivo dos dados1 via TCT

As constatações resultantes da análise da Tabela (3.1.1) encontram suporte nas Figuras (3.1a) e (3.1b) onde são apresentadas as distribuições dos índices de dificuldade e de discriminação do teste para as situações (A) e (B).



(a) Índice de dificuldade incluindo e excluindo o item 2

(b) Índice de discriminação incluindo e excluindo o item 2

Figura 3.1: Índices de dificuldade e de discriminação do teste

3.2 Análise dos dados1 via TRI

Nesta secção, os dados1 são modelados com a finalidade de aferir as habilidades dos alunos na unidade curricular de Bioestatística. Foram utilizados modelos unidimensionais da TRI para dados dicotômicos com recurso ao pacote *ltm* do R. O pacote *ltm* permite utilizar o procedimento iterativo de estimação dos parâmetros de habilidade e dos itens abordando a máxima verossimilhança marginal com recurso ao algoritmo EM, a integração pelo método de Hermite-Gauss e o método de interação Quasi-Newton (*Broyden-Fletcher-Goldfarb-Shanno* - BFGS) com máxima integração. No método BFGS a matriz Hessiana é atualizada iterativamente¹. As estimativas dos parâmetros dos itens (índices de dificuldade, de discriminação e de acerto casual) dos modelos 1PL, 2PL e 3PL, encontram-se na Tabela (A.2) do Apêndice A.

Do ponto de vista estatístico, a escolha de um modelo (neste caso, modelos da TRI) que mais se ajusta aos dados, deve ser de tal modo que o modelo envolva o mínimo de parâmetros possíveis a serem estimados e que explique bem o traço latente da variável resposta. Fazendo jus a esses requisitos, foram utilizados, neste trabalho, os critérios de seleção de

¹Para mais informações sobre o método BFGS veja Nocedal & Wright (2006) e Rao (2009)

modelos *Akaike Information Criterion* (AIC) e *Bayesian Information Criterion* (BIC). Para os modelos em análise, quanto menor for o valor das estatísticas associadas a estes critérios melhor será o ajuste do modelo (Burnham & Anderson, 2004). Adicionalmente, foram analisados os valores da verosimilhança marginalizada e os valores de prova (*p.value*) dos testes de ajustamentos dos modelos considerados aos dados.

Na Tabela (3.2.1) são apresentados, para os modelos 1PL e 2PL, os valores dos critérios de ajustamento mencionados no parágrafo anterior. O valor da verosimilhança marginalizada (log.Lik) mostra um aumento do modelo 1PL para 2PL comprovando a coerência do procedimento computacional pois, com o aumento da complexidade do modelo (mais parâmetros sendo estimados), o valor de log.Lik tende a aumentar. O modelo 2PL tem valores mais baixos do AIC e do BIC em relação ao modelo 1PL conferindo-lhe melhor qualidade de ajuste e ainda, o valor de prova sustenta a existência de melhoria significativa de ajustamento usando modelo 2PL comparativamente a 1PL, isto é, a qualidade do ajuste melhora com o aumento do parâmetro de discriminação.

Modelo	AIC	BIC	log.Lik	p.value
1PL	4685,65	4785,91	-2305,83	
2PL	4610,75	4811,25	-2231,37	<0,001

Tabela 3.2.1: Comparação do ajustamento dos modelos 1PL e 2PL

Por analogia da análise feita no parágrafo anterior, os dados apresentados na Tabela (3.2.2) permitem concluir que o modelo 3PL possui melhor qualidade de ajuste comparativamente ao modelo 1PL.

Modelo	AIC	BIC	log.Lik	p.value
1PL	4685,65	4785,91	-2305,83	
3PL	4652,81	4953,57	-2215,41	<0,001

Tabela 3.2.2: Comparação do ajustamento dos modelos 1PL e 3PL

Finalizando a etapa de seleção de melhor modelo, resta analisar os valores dos critérios de ajustamento para os modelos 2PL e 3PL. Como era de se esperar, verifica-se um aumento do log.Lik do modelo 2PL para o modelo 3PL. Contudo, pelos critérios AIC e BIC, o modelo 2PL é o que possui mais qualidade de ajuste para além de valor de prova ser conclusivo em mostrar a não existência de melhoria significativa na qualidade de ajuste acrescentando ao modelo o parâmetro de acerto ao acaso (ver Tabela (3.2.3)). Doravante, todas as informações apresentadas são resultantes da aplicação do modelo 2PL.

Modelo	AIC	BIC	log.Lik	p.value
2PL	4610,75	4811,25	-2231,37	
3PL	4652,81	4953,57	-2215,41	0,705

Tabela 3.2.3: Comparação do ajustamento dos modelos 2PL e 3PL

Importa lembrar que o modelo 2PL para um item, incorpora os parâmetros de dificuldade e de discriminação e que a probabilidade de um indivíduo, com determinada habilidade, responder corretamente a esse item pode ser representada através da curva característica do item (CCI). Na Figura (3.2) podem-se observar as CCIs dos 37 itens que compõem o teste. Analisando a Figura (3.2) observa-se que, relativamente ao parâmetro dificuldade do item, os itens 1, 3 e 27 são, por um lado, os mais fáceis pois os alunos com menos habilidades (om nível de habilidade abaixo de -2) têm probabilidade de acerto acima de 60%. Entre estes itens, o item 1 é o mais fácil. Por outro lado, os itens 15 e 21 são os mais difíceis visto serem os itens em que os alunos mais hábeis (com nível de habilidade acima de 2) tiveram baixa percentagem de acerto, respetivamente, abaixo dos 65% e 60%. Quanto à discriminação do item, há que destacar dois itens: 2 e 31. Estes itens apresentam um comportamento diferente dos restantes, isto é, dão a entender que quanto mais hábil for um aluno, menor é a probabilidade de acertar no item. Casos como estes merecem muita atenção da parte dos avaliadores desde a revisão até a sua exclusão do instrumento de avaliação. Ainda neste parâmetro, observa-se que o item 18 é o mais discriminativo por apresentar a curva mais íngreme e discrimina melhor os alunos com habilidades entre -1 e 1. Os outros itens com alto poder discriminativos são os itens 9 e 12. Finalmente, tem-se o item 1 com baixa discriminação.

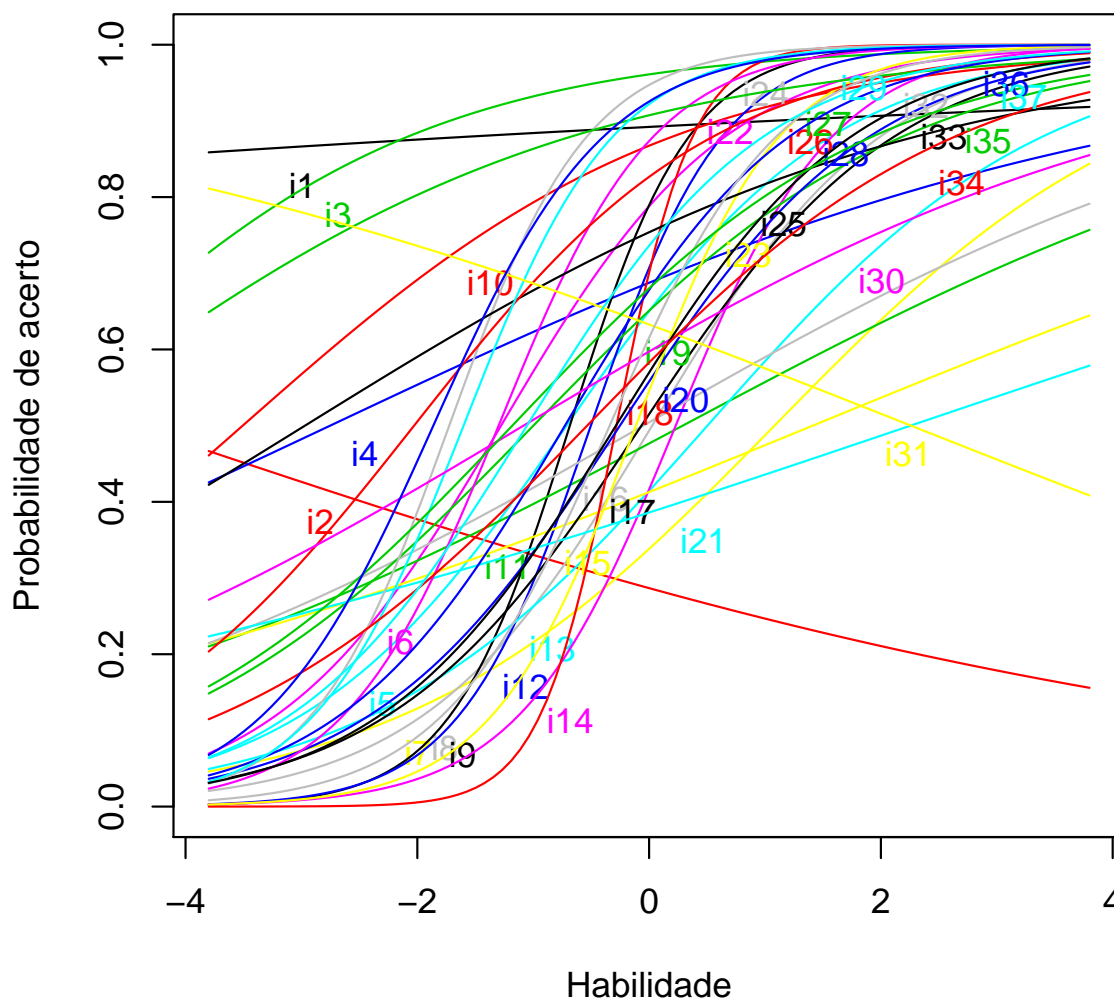


Figura 3.2: Curvas características dos 37 itens do teste

Interessa também saber o grau de contribuição de cada item e do teste em termos de agregação de informação para o teste, tendo em conta a habilidade que se pretende medir. Com recurso à curva de informação do item (CII) e à curva de informação do teste (CIT) é possível fazer essa análise (Figura (3.3) e (3.5)). Efetivamente, na Figura (3.3), o item 18 aparece como aquele que mais contribui com a informação atingindo o ponto máximo da curva em torno de alunos com habilidade -0,2 e, em seguida, surgem os itens 9 e 12. Isto é de se esperar, afinal são os itens constatados como os mais discriminativos. Não basta só observar o nível de contribuição que os itens têm. É também necessário verificar o grupo de alunos

cujas habilidades são melhor estimadas pelos itens. Por exemplo, veja na Figura (3.4) o caso dos itens 14 e 22. Estes itens contribuem, no teste, com quase a mesma quantidade de informação porém, o item 14 fornece melhor estimativas para alunos com habilidades compreendidas entre -3 e 1 ao passo que o item 22, entre -1,5 e 2,5.

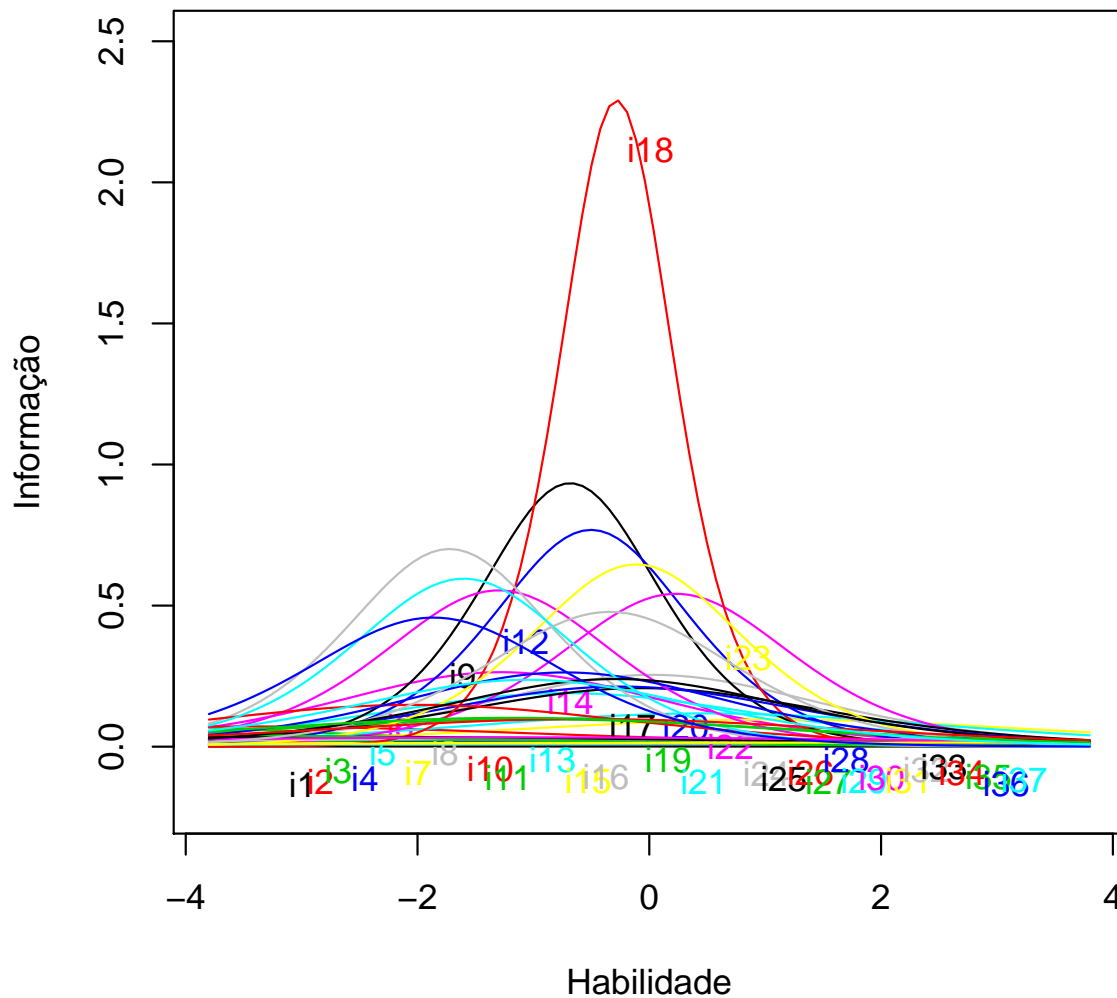


Figura 3.3: Curvas de informação dos 37 itens do teste

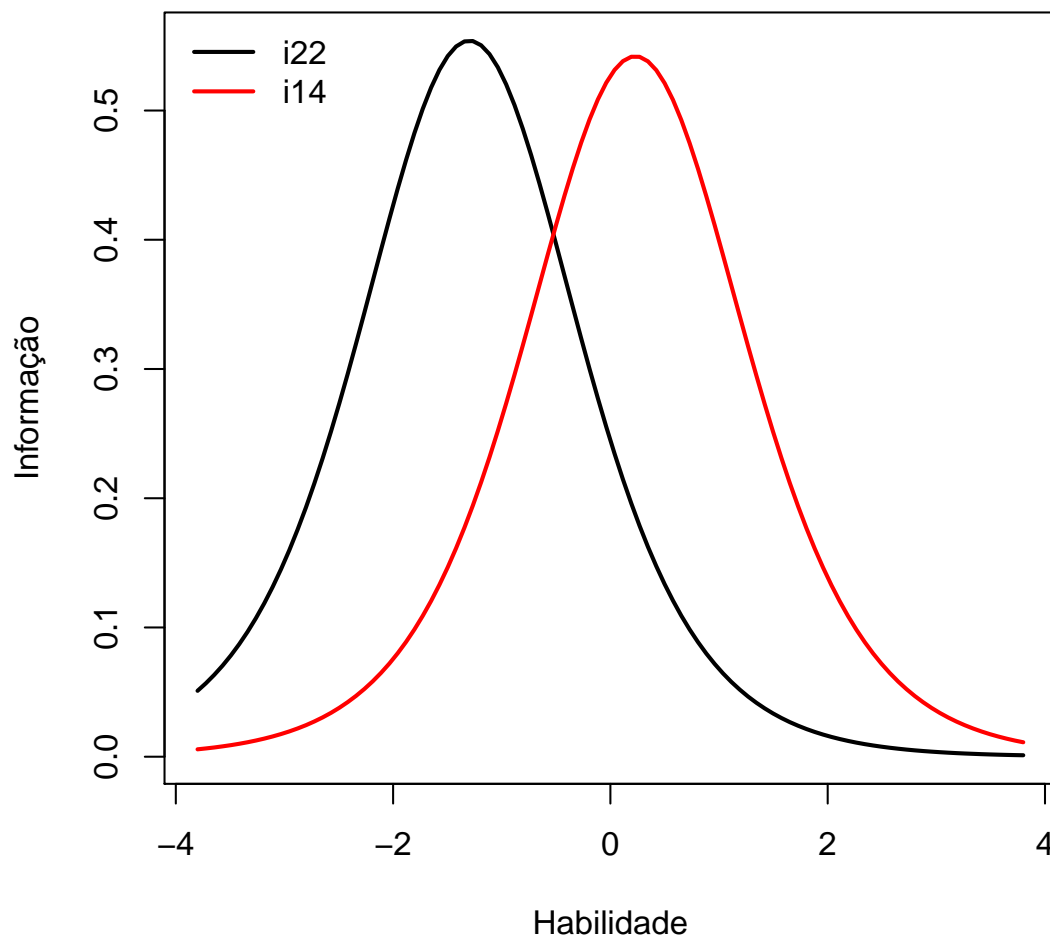


Figura 3.4: Comparação de duas CIIs do teste

A Figura (3.5) apresenta a contribuição do teste na avaliação das habilidades dos alunos, isto é, apresenta a CIT. Não obstante o pico da CIT estar afastado mais para a esquerda do valor central da escala da habilidade, o que seria o ideal pois é onde se situa o maior número de indivíduos, a curva mostra que o teste consegue estimar melhor o desempenho dos alunos que se encontram na escala de habilidade entre -3,0 e 2,0 e que alunos com habilidades acima de 2 não teriam dificuldades em resolver o teste. Neste nível de habilidade, o teste contribui com 83,71% de informação. Com exceção dos itens 2 e 31, a Figura (3.2) mostra claramente que alunos com habilidades acima de 2 têm probabilidades mínimas de 50% de responderem corretamente às questões do teste.

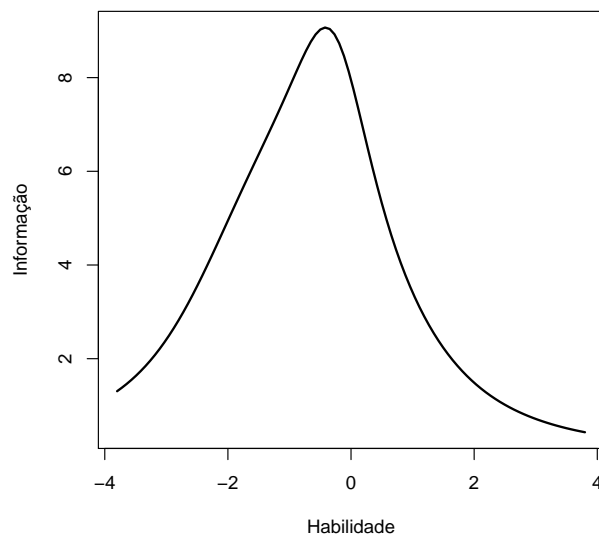


Figura 3.5: Curva de informação do teste

Como estão, então, os alunos distribuídos com base nas suas habilidades? A partir da Figura (3.6) observa-se que a maioria dos alunos têm habilidade entre -0,5 e 0, seguido do grupo entre 0 e 1. Nota-se ainda que existe apenas 1 aluno com habilidade abaixo de -2 e que não existem alunos com habilidade acima de 2. Para mais detalhes, a Tabela (A.3) do Apêndice A, fornece as estimativas das habilidades para cada um dos 111 alunos.

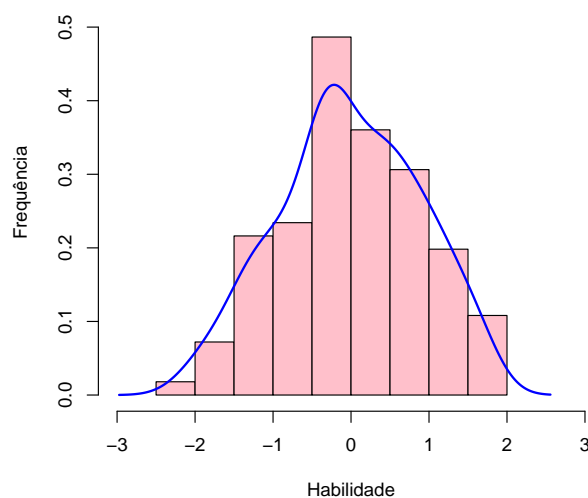


Figura 3.6: Distribuição dos alunos na escala das habilidades

3.3 Análise dos dados2

Os dados2 apresentam avaliações de natureza ordinais e portanto, para analisar o nível de concordância dos professores inquiridos quanto à validade e consistência do teste aplicado aos alunos descritos no início deste capítulo, recorreu-se à medida conhecida como coeficiente de concordância W de Kendall. De acordo com Siegel (1975), W de Kendall é um método não-paramétrico que permite calcular a concordância entre três ou mais avaliadores da classificação que eles dão ou fazem a um instrumento de avaliação (no contexto desta dissertação), de acordo com uma característica (critério) específica. O seu valor varia entre 0 e 1. Os resultados de W de Kendall apresentados a seguir, levam em conta a classificação que os 6 avaliadores (professores inquiridos) deram aos 37 itens que compõem o teste em função de 4 critérios: (i) clareza de linguagem do enunciado e das opções de respostas, (ii) grau de dificuldade da questão, (iii) adequação da questão ao conteúdo curricular e (iv) capacidade de a questão discriminar os bons alunos dos menos bons. Uma outra avaliação paralela foi feita com base nos mesmos critérios mas desta vez, para os 5 momentos de avaliação que perfazem o teste.

Os valores de W de Kendall têm interpretação idêntica aos de Kappa de Cohen (Cohen, 1960). A estatística Kappa é mais utilizada para medir a concordância das avaliações nominais. Kappa assume valores entre -1 e 1 porém, valores negativos sugerem total discordância entre os avaliadores e não tem uma interpretação específica em termos de grau de discordância. Valores negativos refletem, muitas das vezes, níveis de concordância inferiores aos esperados pelo acaso. De uma forma geral, enquanto os valores de Kappa muito próximos de 0 sugerem fraca concordância a qual se deve exclusivamente ao acaso, valores muito próximos de 1, sugerem o oposto (Fonseca et al, 2007). Landis & Koch (1977) propõem, através da Tabela (3.3.1), interpretação para diferentes valores de Kappa:

Valor de Kappa	Nível de Concordância
<0.00	Má
0.00 - 0.20	Fraca
0.21 - 0.40	Razoável
0.41 - 0.60	Moderada
0.61 - 0.80	Substancial
0.81 - 1.00	Quase perfeita

Tabela 3.3.1: Nível de concordância

Comparando as interpretações para valores de Kappa segundo Landis & Koch (1977) e autores como Fleiss (1981) e Altman (1991), é recomendável que Kappa tenha valores acima de 0.70 para se concluir que os avaliadores utilizaram o mesmo padrão de avaliação.

Os 6 avaliadores considerados para o teste foram vinculados a universidades públicas de Portugal e estão ligados às áreas da Estatística. Todos os 6 têm nível académico de doutorado. Relativamente ao tempo de serviço, a metade deles tem tempo de serviço superior a 20 anos, 2 entre 11 e 20 anos e 1 entre 6 e 10 anos. Esta descrição qualifica os avaliadores como sendo apropriados para a tarefa.

Na Tabela (3.3.2) são apresentados os valores do coeficiente de concordância W de Kendall para os 37 itens do teste e, na Tabela (3.3.3), para os 5 momentos de avaliação.

Item	Valor de W	Item	Valor de W	Item	Valor de W
i1	0.503	i14	0.713	i26	0.5
i2	0.261	i15	0.446	i27	0.81
i3	0.934	i16	0.489	i28	0.533
i4	0.701	i17	0.264	i29	0.343
i5	0.33	i18	0.346	i30	0.486
i6	0.714	i19	0.536	i31	0.779
i7	0.626	i20	0.557	i32	0.609
i8	0.153	i21	0.451	i33	0.765
i9	0.321	i22	0.51	i34	0.84
i10	0.648	i23	0.765	i35	0.815
i11	0.552	i24	0.667	i36	0.447
i12	0.573	i25	0.636	i37	0.358
i13	0.614				

Tabela 3.3.2: Nível de concordância para cada item

Momento de Avaliação	Valor de W
1	0.382
2	0.479
3	0.443
4	0.167
5	0.329

Tabela 3.3.3: Nível de concordância para cada momento de avaliação

Em relação aos itens avaliados isoladamente, verificam-se todas as situações descritas na Tabela 3.3.1, com exceção de nível de concordância mau. Os itens 2, 5, 8, 9, 18, 29 e 37 têm níveis de concordância abaixo de moderada sendo o item 8 o que apresenta a concordância fraca. O nível de concordância quase perfeita verifica-se aos itens 3, 27, 34 e 35. Os restantes itens (a maioria), apresentam o nível de concordância entre moderada e substancial. De um modo geral, os valores do coeficiente de concordância W revelam a disparidade

de padrões adotados pelos avaliadores na avaliação dos itens. Analisando as pontuações atribuídas à cada item, foi notória a divergência dos avaliadores nos critérios de grau de dificuldade do item e adequação do item ao conteúdo curricular. Alguns comentários diziam que a dificuldade de um item depende de fatores como material de apoio autorizado durante a realização do teste ou mesmo das fichas de exercícios discutidas nas aulas. Outros comentários condicionaram a avaliação objetiva do item, no critério de adequação do item ao conteúdo curricular, à consulta do programa de disciplina. Portanto, julgamos que o baixo nível de concordância verificada na maioria dos itens esteja ligado a esses condicionalismos. Analogamente, explicam-se os baixos valores do coeficiente de concordância W na avaliação dos momentos de avaliação.

Depois de submetidos a uma avaliação, os alunos têm por hábito identificarem as questões que acharam mais fácil ou mais difícil e, algumas vezes, os docentes têm noção da existência desse tipo de questões. Achamos interessante verificar se os itens considerados fáceis e difíceis pelos alunos (a partir dos resultados do teste) são os mesmos, quando comparados com a classificação feita pelos avaliadores. Os itens tidos como os mais fáceis e mais difíceis, na percepção dos avaliadores, constam da Tabela (3.3.4). Seja MA - momento de avaliação e, Av - avaliador:

MA	Item mais fácil						Item mais difícil					
	Av.1	Av.2	Av.3	Av.4	Av.5	Av.6	Av.1	Av.2	Av.3	Av.4	Av.5	Av.6
1	3	3	1	3	3	3	2	5	3	2	4	5
2	7	6	6	13	6	9	11	12	9	11	8	8
3	20	20	15	17	21	14	21	21	16	19	15	17
4	23	24	23	24	24	24	27	27	27	22	26	29
5	30	36	33	31	33	36	36	35	36	37	36	35
Teste	23	3	33	3	3	3	36	21	36	37	36	29

Tabela 3.3.4: Itens mais fáceis e mais difíceis na percepção dos avaliadores

Nota-se que os itens considerados como sendo os fáceis por alguns avaliadores, são considerados como sendo os mais difíceis por outros. Veja o caso do item 3 em MA1 onde cinco dos

avaliadores consideram-no fácil porém, um deles acha o contrário. Relativamente ao teste, analisando pela opinião da maioria, pode-se dizer que o item 3 foi o mais fácil ao passo que o item 36, o mais difícil. Confrontando a opinião dos avaliadores com o resultado dos alunos (ver Tabela (3.3.5)) é, visivelmente, notória a discrepância com exceção de alguns casos isolados. Verifica-se algum alinhamento na questão mais fácil para o MA1 e para o teste. O mesmo não se pode dizer em relação às questões difíceis. Aliás, no MA5, o item 36 é considerado pelos avaliadores como sendo o mais difícil contudo, os resultados do teste provaram o contrário.

MA	Item mais fácil	Item mais difícil
1	3	2
2	10	7
3	19	21
4	27	23
5	36	33
Teste	3	2

Tabela 3.3.5: Itens mais fáceis e mais difíceis para os alunos

Era provável que, em algum momento, não houvesse convergência entre o resultado dos alunos nos testes e a opinião dos avaliadores quanto aos itens fáceis e itens difíceis. Aparentemente o parâmetro de dificuldade é fácil de identificar e avaliar porém, conforme vimos na aplicação da TRI, o nível de habilidade do aluno é determinante para se ser conclusivo. Para além do parâmetro de dificuldade, um item pode integrar o parâmetro de discriminação e de acerto casual e, em conjunto, todos dependem da habilidade do aluno. Portanto, ao se elaborar um teste, os itens devem satisfazer os objetivos do avaliador, ao mesmo tempo que se leva em conta o nível de habilidade dos alunos em relação aos conteúdos ou níveis de ensino a avaliar. Para o efeito, sugere-se a criação de um banco de itens calibrados a partir do qual se possa seleccionar os itens para compor um teste que melhor se ajuste ao nível dos alunos. Entende-se por itens calibrados aqueles itens cujos parâmetros já são conhecidos (depois de testados a partir de um número significativo de uma população) possibilitando a estimação do nível de habilidade dos alunos através de um conjunto representativo de itens. Para Andrade et al (2000), a criação de um banco de itens é contínuo e, portanto, sempre que houver

necessidade de acrescentar novos itens, estes devem ser calibrados na mesma escala em que estão os outros itens do banco. Naturalmente que não se deve por de lado a quantidade de informação dos itens para diferentes níveis de habilidade.

Capítulo 4

Exercícios parametrizados

Os exercícios parametrizados têm sido bastante utilizados por professores e investigadores ligados à área do Ensino. A maioria desses profissionais têm usado ambientes *web* a partir dos quais vários exercícios são gerados permitindo que os alunos façam autoestudo e auto-avaliação numa mesma unidade temática e que os professores tenham um banco de dados de exercícios semelhantes quer pelos conteúdos quer pelo nível de dificuldades ou ainda pelo nível de discriminação (PmatE, 2010; MEGUA, 2010). Contudo, estes ambientes *web*, na sua maioria, fornecem aos utentes apenas a resposta correta dos exercícios sem no entanto apresentarem a resolução, caso do PmatE - Projecto Matemática Ensino da Universidade de Aveiro - Oliveira et al (2014). Neste trabalho, para além da elaboração de exercícios parametrizados e apresentação da resposta correta, são apresentadas as propostas de resoluções que conduzem à resposta correta bem como as possíveis causas que contribuem para obtenção de respostas erradas dos exercícios. Um projeto similar com fins pedagógicos designado por MEGUA (MEGUA, 2010), foi criado para auxiliar os alunos a serem autodidatas na medida em que, a partir dos exercícios disponibilizados na base de dados da plataforma, eles podem resolver diferentes exercícios sob o mesmo conteúdo e fazer a verificação dos resultados com base nas propostas de resolução disponibilizadas. Os professores são, também, beneficiados pois contribui na redução da atividade periódica de elaboração de exercícios sobre mesmos conteúdos, flexibiliza a produção de material didático e de apoio às aulas e à avaliação e pode, inclusive, reduzir o tempo necessário para esclarecer possíveis dúvidas aos alunos (Cruz et al, 2013).

Os exercícios apresentados neste trabalho foram desenvolvidos a partir do ambiente de trabalho do *SageMathCloud* (um *software* matemático livre e de código aberto, desenvolvido sob a licença GPL por uma comunidade de programadores e matemáticos, que busca ser uma alternativa para os principais sistemas proprietários de *software* matemático como o Magma, Maple, Mathematica e Matlab) que integra, dentre várias funções, a linguagem $\text{L}^{\text{A}}\text{T}_{\text{E}}\text{X}$ (para efeitos de produção de texto matemático) e do R (para gerar valores aleatórios e calcular alguns indicadores numéricos). Uma das vantagens desta plataforma é que qualquer outro pesquisador que receba páginas de *SageMathCloud* é capaz de as partilhar e manipular sem a necessidade de aquisição de *software* proprietário (Stein, 2011). Isto foi notório aquando da utilização da plataforma para a produção deste trabalho. Essa funcionalidade tornou possível o acesso, a manipulação e a interação entre intervenientes sem que estivessem juntos presencialmente.

4.1 Conteúdos tratados

Os exercícios abordam, por opção, o tópico das variáveis aleatórias discretas. A cada exercício modelo estão associadas 4 respostas possíveis. A elaboração da resposta correta leva em conta todos os procedimentos válidos para a sua obtenção. As respostas erradas resultam das constatações de alguns erros cometidos pelos alunos verificadas no decurso da atividade docente do autor desta dissertação enquanto professor de Matemática em Moçambique, nomeadamente: interpretação errada ao nível do acontecimento em causa, dificuldades em distinguir a função massa de probabilidade da função acumulada de probabilidade, aplicação errada de propriedades da esperança e da variância, entre outros. A seguir são apresentados, resumidamente, alguns conceitos sobre variáveis aleatórias discretas baseando-se nas obras de Fonseca (2001), Casella & Berger (2002) e Murteira et al (2015).

Definição 4.1: A função $f(x)$ chama-se **função massa de probabilidade** (*probability mass function*) ou **função probabilidade** da variável aleatória (v.a.) discreta X se e só se:

$$f(x) = \begin{cases} P(X = x), & \text{se } x = x_i \\ 0, & \text{se } x \neq x_i \end{cases} \quad i = 1, 2, \dots, n \quad (4.1)$$

Definição 4.2: Designa-se por **função acumulada de probabilidade** (*cumulative distribution function*) ou **função de distribuição** de da v.a. X à função real de variável real F , definida por

$$F(x) = P(X \leq x), \forall x \in \mathbb{R}. \quad (4.2)$$

Toda função de distribuição F satisfaz as seguintes propriedades:

- $0 \leq F(x) \leq 1$;
- $F(-\infty) = \lim_{x \rightarrow -\infty} F(x) = 0$ e $F(\infty) = \lim_{x \rightarrow \infty} F(x) = 1$;
- $F(x)$ é não decrescente: $\Delta x > 0 \Rightarrow F(x) \leq F(x + \Delta x)$;
- $P(a < X \leq b) = F(b) - F(a)$, $\forall a, b \in \mathbb{R}$ desde que $a < b$;
- F é contínua à direita: $F(a^+) = \lim_{x \rightarrow a^+} F(x) = F(a)$;
- $P(X = a) = F(a) - F(a^-)$, com $F(a^-) = \lim_{x \rightarrow a^-} F(x)$.

Definição 4.3: Seja X uma v.a. discreta que assume diferentes valores reais $x_1, x_2, \dots, x_n, \dots$. O **valor esperado** (a média ou a esperança) de X , denotado por $E(X)$ ou simplesmente μ , é dado por

$$E(X) = \sum_{i=1}^{\infty} x_i f(x_i). \quad (4.3)$$

Este valor tem como uma das propriedades a linearidade do operador:

$$E(aX \pm b) = E(aX) \pm E(b) = aE(X) \pm b \quad (4.4)$$

onde a e b são constantes reais.

Definição 4.4: Seja X uma v.a. discreta que assume diferentes valores $x_1, x_2, \dots, x_n, \dots$. A **variância** de X , simbolicamente representada por $V(X)$ ou simplesmente σ^2 , é definida pela expressão

$$V(X) = E[X - E(X)]^2 \quad \text{ou} \quad V(X) = E(X^2) - [E(X)]^2. \quad (4.5)$$

Para quaisquer constantes reais a e b , é válida a propriedade

$$V(aX + b) = a^2 V(X). \quad (4.6)$$

A parametrização que implementamos nos exercícios que a seguir propomos, consistiu em criar um algoritmo que permitisse gerar aleatoriamente, dentro de um conjunto definido, probabilidades associadas aos valores que a variável aleatória discreta assume. Tendo em conta as Definições 4.1 e 4.2, foi aplicado o seguinte algoritmo para a obtenção dessas probabilidades, partindo de valores iniciais para n , a e b .

1º Passo: gerar n valores aleatórios inteiros $(x_1, x_2, \dots, x_i, \dots, x_n)$ compreendidos entre a e b , por exemplo e que funcionam como pesos de cada observação distinta x_i que irá fazer parte dos enunciados aleatórios;

2º Passo: determinar a probabilidade para ocorrência de cada um desses valores x_i usando a ponderação, exceto para o último isto é,

$$f(x_i) = \frac{x_i}{\sum_{i=1}^n x_i}, \quad i = 1, 2, \dots, n-1; \quad (4.7)$$

3º Passo: determinar o valor da n -ésima probabilidade de ocorrência de x_n , isto é,

$$f(x_n) = 1 - \sum_{i=1}^{n-1} f(x_i). \quad (4.8)$$

O terceiro passo visa garantir, efetiva e numericamente, que a soma de todas as probabilidade seja igual a um. Este procedimento é importante para resolver problemas causados por arredondamentos em (4.7).

4.2 Descrição e proposta de resolução de exercícios

Quatro exercícios modelo com 4 opções de escolhas múltipla (com apenas uma opção certa) são apresentados nesta secção. Para cada um dos exercícios é apresentado o enunciado, a descrição, a proposta de resolução correta e as possíveis causas das respostas erradas que constam nas opções não corretas.

4.2.1 Exercício 1

Enunciado

Um agricultor tem os seguintes valores para as estimativas das probabilidades correspondentes aos dias necessários para terminar determinada sementeira:

Número de dias	1	2	3	4	5
Probabilidades	0.05	0.20	0.35	0.30	0.10

A probabilidade de que uma sementeira escolhida aleatoriamente leve não mais de quatro dias é:

- (a) 0.90
- (b) 0.10
- (c) 0.60
- (d) 0.40

Parametrização

O enunciado é gerado automaticamente sempre que se executa o comando para a respetiva produção. Nesse processo, os valores das probabilidades são obtidos aleatoriamente por substituição de parâmetros definidos por valores numéricos e seguidamente são calculadas as opções de respostas. A v.a. número de dias para terminar uma sementeira toma somente 5 valores e, portanto, são gerados aleatoriamente 5 números inteiros entre $a = 1$ e $b = 20$: x_1 , x_2 , x_3 , x_4 e x_5 (1º Passo). A seguir (2º Passo) são calculadas as probabilidades correspondentes: p_1 , p_2 , p_3 e p_4 exceto a última (p_5), isto é, $p_1 = \frac{x_1}{S}$, $p_2 = \frac{x_2}{S}$, $p_3 = \frac{x_3}{S}$ e $p_4 = \frac{x_4}{S}$, onde $S = x_1 + x_2 + x_3 + x_4 + x_5$. Finalmente (3º Passo) calcula-se o valor da última probabilidade: $p_5 = 1 - (p_1 + p_2 + p_3 + p_4)$.

As opções de respostas quer a correta quer as incorretas são, igualmente, calculadas automaticamente utilizando os valores das probabilidades obtidas aleatoriamente tendo em conta procedimentos pré-definidos. Esses procedimentos são descritos na proposta de resolução e

na justificação do porquê da resposta ser incorreta (falsa).

Descrição e proposta de resolução

Neste exercício espera-se que o aluno, a partir de uma tabela que traduz a função massa de probabilidade, calcule as probabilidades de ocorrência de qualquer evento. A variável aleatória $X = \text{“Número de dias para terminar uma sementeira”}$ é discreta, as probabilidades associadas aos valores possíveis da variável são positivas e o seu somatório é igual à unidade. Verificadas essas condições, conclui-se que se trata de uma função massa de probabilidade. O pedido “a probabilidade de que uma sementeira escolhida aleatoriamente leve não mais de quatro dias” equivale a dizer “a probabilidade de que uma sementeira escolhida aleatoriamente leve um número de dias inferior ou igual a quatro”, isto é,

$$P(X \leq 4).$$

Lembre-se que para uma v.a. discreta, a função massa de probabilidade é dada pela expressão (4.1). Para o problema em causa, X assume números inteiros de 1 a 5 e os números iguais ou inferiores a 4 são 1, 2, 3 e 4. Assim, com base na tabela da função massa de probabilidade, dada por:

	x_1	x_2	x_3	x_4	x_5
x_i	1	2	3	4	5
$P(X = x_i)$	0.05	0.20	0.35	0.30	0.10

tem-se:

$$\begin{aligned}
 P(X \leq 4) &= P(X = x_1) + P(X = x_2) + P(X = x_3) + P(X = x_4) \\
 &= P(X = 1) + P(X = 2) + P(X = 3) + P(X = 4) \\
 &= 0.05 + 0.20 + 0.35 + 0.30 \\
 &= 0.90.
 \end{aligned}$$

Portanto:

(a) Verdadeira.

- (b) É falsa porque considera a probabilidade correspondente à sementeira levar um número de dias superior 4, isto é, $P(X > 4)$.
- (c) É falsa porque considera a probabilidade da sementeira terminar em um número dias estritamente inferiores a 4, isto é, $P(X < 4)$.
- (d) É falsa porque considera a probabilidade da sementeira terminar em um número de dias igual ou superior a 4, isto é, $P(X \geq 4)$.

As opções erradas propostas neste exercício têm um aspecto em comum: má interpretação do evento de interesse.

4.2.2 Exercício 2

Enunciado

A distribuição de probabilidades relativa ao número de carros vendidos numa determinada semana, é dada pela seguinte tabela:

Número de carros vendidos ser menor ou igual	0	1	2	3	4	5
Probabilidades	0.10	0.30	0.65	0.83	0.95	1.0

A probabilidade de que, numa determinada semana, sejam vendidos exatamente três carros é:

- (a) 0.18
- (b) 0.83
- (c) 0.95
- (d) 0.55

Parametrização

Neste exercício, apesar de a v.a. tomar 6 valores, a parametrização segue a lógica utilizada no Exercício 1. Porém, observa-se que as probabilidades são acumuladas, isto é, estão associadas a um ou mais valores da v.a. Assim, depois de calculadas as probabilidades pontuais p_1, p_2, p_3, p_4, p_5 e p_6 são obtidas as probabilidades acumuladas P_1, P_2, P_3, P_4, P_5 e P_6 tais que: $P_1 = p_1, P_2 = P_1 + p_2, P_3 = P_2 + p_3, P_4 = P_3 + p_4, P_5 = P_4 + p_5$ e $P_6 = P_5 + p_6$. A partir daí são obtidas as opções de respostas conforme se observa a seguir.

Descrição e proposta de resolução

Neste exercício espera-se que o aluno, a partir de uma tabela que traduz a função acumulada de probabilidade, calcule as probabilidades de ocorrência de qualquer evento associado a essa v.a. A v.a. $X = \text{“Número de carros vendidos por semana”}$ é discreta, as probabilidades associadas aos valores possíveis da variável são positivas e, na tabela dada, o somatório das probabilidades não é igual à unidade. Analisando na tabela as probabilidades associadas aos valores acumulados da variável aleatória X , nota-se que estas são crescentes e que a última probabilidade é exatamente igual a 1. Estas condições caracterizam a função acumulada de probabilidade. Na sequência, apresenta-se a seguir, a proposta de resolução para “a probabilidade de, em uma determinada semana, vender exatamente três carros”, que é o mesmo que determinar $P(X = 3)$.

Levando em consideração as características acima, $P(X = 3)$ pode ser encontrada aplicando a propriedade

$$P(X = a) = F(a) - F(a^-), \quad \text{com} \quad F(a^-) = \lim_{x \rightarrow a^-} F(x).$$

Assim,

$$P(X = 3) = F(3) - F(3^-) = F(3) - F(2) = 0.83 - 0.65 = 0.18.$$

Para as opções apresentadas temos:

(a) Verdadeira.

- (b) É falsa porque considera a probabilidade correspondente à venda de 3 carros, conforme a tabela do enunciado, como se de uma função massa de probabilidade se tratasse, isto é, determina $F(3)$ e não $P(X = 3)$.
- (c) É falsa porque considera a soma dos valores das probabilidades apresentadas na tabela como se correspondesse a uma tabela da função massa de probabilidade e toma os valores correspondentes a venda de um ou dois carros, isto é, faz $F(1) + F(2)$.
- (d) É falsa porque, apesar de encontrar a função massa de probabilidade, considera a soma dos valores das probabilidades correspondentes à venda de um ou dois carros, isto é, faz $P(X = 1) + P(X = 2)$.

4.2.3 Exercício 3

Enunciado

Considere o enunciado do Exercício 2. Sabe-se que, por semana, o vendedor recebe um salário fixo de 300 euros mais 150 euros por cada carro vendido. Para a semana seguinte, o salário esperado e a variância do salário semanal do vendedor são, respetivamente:

- (a) 765 e 70425
- (b) 1080 e 281700
- (c) 1827 e -1377729
- (d) 1950 e 70725

Parametrização

Tratando-se do mesmo enunciado do Exercício 2, mantém-se o processo de parametrização aplicado.

Descrição e proposta de resolução

Neste exercício, espera-se que o aluno, analisando os valores crescentes das probabilidades para valores acumulados da variável observada, compreenda que a tabela traduz a função acumulada de probabilidade e, a partir desta, encontre a função massa de probabilidade por forma a calcular a esperança e a variância da variável que traduz o salário semanal do vendedor.

Denotando por X a v.a. relativa ao “número de carros vendidos por semana” calculam-se, a seguir, as probabilidades associadas aos valores de X . O procedimento a aplicar é análogo ao descrito na resolução do Exercício 2, sendo aqui para todos os valores possíveis de X . Com efeito:

$$P(X = 0) = F(0) - F(0^-) = 0.16 - 0 = 0.16$$

$$P(X = 1) = F(1) - F(1^-) = 0.23 - 0.16 = 0.07$$

$$P(X = 2) = F(2) - F(2^-) = 0.32 - 0.23 = 0.09$$

$$P(X = 3) = F(3) - F(3^-) = 0.45 - 0.32 = 0.13$$

$$P(X = 4) = F(4) - F(4^-) = 0.74 - 0.45 = 0.29$$

$$P(X = 5) = F(5) - F(5^-) = 1 - 0.74 = 0.26$$

Assim, a tabela da função massa de probabilidade é

X	0	1	2	3	4	5
$P(X = x)$	0.16	0.07	0.09	0.13	0.29	0.26

Denotando a *comissão de venda* por c e o *salário fixo* por a , o salário esperado será dado por

$$E(cX + a)$$

e, pela propriedade de linearidade do operador esperança, resulta que

$$E(cX + a) = E(cX) + E(a) = cE(X) + a.$$

Sendo conhecidos os valores de c e de a , resta calcular a esperança $E(X)$ aplicando a fórmula (4.3):

$$\begin{aligned} E(X) &= \sum_{i=1}^6 x_i f(x_i) \\ &= \sum_{i=0}^5 i \times P(X = i) \\ &= 0 \times 0.16 + 1 \times 0.07 + 2 \times 0.09 + 3 \times 0.13 + 4 \times 0.29 + 5 \times 0.26 \\ &= 3.1, \end{aligned}$$

consequentemente

$$E(cX + a) = 150 \times 3.1 + 300 = 765.$$

Relativamente à variância, podemos obter o seu valor com base na propriedade (4.6) e ainda da fórmula (4.5).

Ora,

$$\begin{aligned} E(X^2) &= \sum_{i=1}^6 x_i^2 f(x_i) \\ &= \sum_{i=0}^5 i^2 \times P(X = i) \\ &= 0^2 \times 0.16 + 1^2 \times 0.07 + 2^2 \times 0.09 + 3^2 \times 0.13 + 4^2 \times 0.29 + 5^2 \times 0.26 \\ &= 12.74. \end{aligned}$$

Logo,

$$V(X) = 12.74 - 3.1^2 = 3.13.$$

Finalmente é calculada a variância do salário esperado dada por:

$$V(cX + a) = c^2 V(X) = 150^2 \times 3.13 = 70425.$$

Portanto, para as opções sugeridas conclui-se:

- (a) Verdadeira.
- (b) Falsa. Esta opção resulta da troca dos valores dos parâmetros a e c , isto é, $E(300X + 150)$ e $V(300X + 150)$.
- (c) Falsa. Esta opção resulta de usar as probabilidades de uma função acumulada de probabilidades para efetuar os cálculos da esperança e da variância pedidas.
- (d) Falsa. Esta opção resulta de tomar o cálculo da esperança a partir da média aritmética dos valores da variável X sem ter em conta as probabilidades a eles associados e aplica erradamente a propriedade da variância fazendo $V(150X + 300) = 150^2V(X) + 300$.

4.2.4 Exercício 4

Enunciado

Os relatórios de uma agência bancária mostram que, em média, são atendidos 75 clientes por dia. Por forma a aumentar a eficiência dos seus funcionários, o administrador da agência bancária criou um sistema que oferece a cada funcionário um prémio de 750 euros por cada cliente extra atendido acima de 76 clientes por dia. O ganho operacional da agência é de 650 euros, para cada cliente extra atendido acima de 75 clientes por dia. Para cada um dos casos (oferecer prémio ou ter um ganho), o teto máximo é de 80 clientes por dia. As probabilidades de atendimento são:

Número de clientes	75	76	77	78	79	80
Probabilidades	0.23	0.19	0.06	0.2	0.1	0.22

Caso o sistema seja implementado, a agência espera ter um lucro (valor positivo) ou prejuízo (valor negativo) de:

- (a) 336.5
- (b) 2250
- (c) 2796.5

(d) -336.5

Parametrização

Neste exercício, o processo de parametrização da tabela da função massa de probabilidade dada no enunciado é análogo ao aplicado ao Exercício 1.

Descrição e proposta de resolução

A ideia do exercício é obter o lucro (L) esperado pela agência com a implementação de um novo sistema, a partir da informação dos números de clientes que os funcionários da agência atendem por dia, os custos envolvidos e os respectivos ganhos. Assim sendo, o aluno precisa verificar que a tabela dada corresponde a uma função massa de probabilidade e seguidamente substituir os valores da variável “número de clientes atendidos por dia” pelos prêmios (custos) e pelos ganhos (receitas), no contexto do problema. O aluno deve observar que a nova variável (L) resulta da diferença entre a receita (R) e o custo (C): $L = R - C$. Sendo L uma v.a. discreta, o lucro esperado é obtido aplicando a fórmula (4.3) para o cálculo da esperança:

$$E(L) = \sum_{i=1}^6 l_i f(l_i) \quad \text{onde} \quad l_i = r_i - c_i.$$

Por opção, os cálculos preliminares são apresentados na tabela a seguir:

X	75	76	77	78	79	80	Total
R	0	650	1300	1950	2600	3250	-
C	0	0	750	1500	2250	3000	-
L	0	650	550	450	350	250	-
$P(L = l)$	0.23	0.19	0.06	0.2	0.1	0.22	1.0

$$E(L) = 0 \times 0.23 + 650 \times 0.19 + 550 \times 0.06 + 450 \times 0.2 + 350 \times 0.1 + 250 \times 0.22 = 336.5$$

Portanto:

(a) Verdadeira.

- (b) Falsa. Calcula o lucro esperado somente a partir da relação $L = R - C$, tomando as probabilidades dada na tabela do enunciado e portanto ignorando completamente as corretas probabilidades associadas.
- (c) Falsa. Esta opção resulta de considerar o lucro como a soma entre a receita e o custo ($L = R + C$).
- (d) Falsa. Esta opção resulta de considerar o lucro como a diferença entre o custo e a receita ($L = C - R$).

Considerações Finais

No processo de ensino e de aprendizagem a avaliação (teste) é, indubitavelmente, um instrumento fundamental para medir, qualitativa e quantitativamente, o nível de aprendizagem alcançado pelos alunos bem como verificar até que ponto os objetivos programáticos para um determinado conteúdo curricular, unidade temática, capítulo, programa ou ciclo acadêmico foram alcançados. Um teste deve conter itens cujas características (parâmetros) dêem ao professor garantias sólidas de que esteja a medir, individualmente, o nível de habilidade dos alunos. Nesta dissertação foram apresentadas, descritas e discutidas teorias que auxiliam os professores na avaliação de testes: a Teoria Clássica de Testes (TCT) e a Teoria de Resposta ao Item (TRI). Na sequência, fez-se aplicação da TCT e da TRI a dados resultantes do processo de avaliação de alunos na unidade curricular de Bioestatística. Apresentámos ainda, em jeito de proposta, quatro exercícios parametrizados. A ideia foi de mostrar como a criação e uso deste tipo de exercício pode facilitar a tarefa do professor de elaborar, periodicamente, exercícios para os mesmos conteúdos.

A aplicação da TCT aos dados obtidos através do teste feito pelos alunos, permitiu identificar itens fáceis e difíceis tendo em conta a proporção de acertos, permitiu verificar se todos os itens medem a mesma habilidade e até que ponto os itens estão correlacionados. Estes indicadores foram fundamentais para descrever as características do teste. Com efeito, verificou-se que o teste era composto de questões com todos os níveis de dificuldade (na sua maioria de dificuldade média) e de discriminação (na sua maioria boa). Desta análise conclui-se que o item 2 não era adequado para medir a habilidade dos alunos devendo ser removido e que com esta ação, melhorar-se-ia, globalmente, a consistência interna do teste.

As características peculiares da TRI foram de extrema importância na análise de cada item

que compõe o teste. Assim sendo, depois de selecionar o modelo que apresentava melhor ajuste (modelo logístico de 2 parâmetros), foram estimados, para cada item, os parâmetros de dificuldade, de discriminação e as habilidades dos alunos. As curvas características de itens e as curvas de informação de itens contribuíram na identificação de maus itens, itens que discriminam os alunos mais proficientes dos menos proficientes e itens que contribuem com mais informação na estimação das habilidades dos alunos. Para além do item 2 identificado como mau a partir da TCT, a TRI identificou o item 31. Estes itens apresentaram um comportamento diferente dos demais itens ao darem a entender que alunos mais proficientes têm menos probabilidades de os responderem corretamente. Com exceção destes itens, o teste demonstrou ser um instrumento com boa precisão e válido para alunos na escala de habilidades de -3,0 até 2,0.

Os resultados relativos à avaliação do teste mostraram, para a maioria dos itens num total de 28 em 37, o nível de concordância abaixo de 0,7. Esses valores permitem concluir que os professores (avaliadores) não utilizaram o mesmo padrão para avaliação. Algumas hipóteses que levantamos foram a questão de relatividade na aplicação e interpretação de alguns critérios de avaliação (caso do grau de dificuldade do item) e a falta de informação que permitisse sustentar o posicionamento dos avaliadores (caso do programa da disciplina e plano curricular). No entanto, os mesmos resultados deixaram claro que nem sempre a forma como o professor analisa um teste corresponde à expectativa dos alunos podendo estar acima ou abaixo das suas habilidades.

A TCT e a TRI contribuíram, significativamente, na avaliação do teste em geral e dos itens, em particular. Conseguimos identificar, a partir dos resultados dos alunos, a consistência, o grau de dificuldade e de discriminação de cada item. Conseguimos verificar o contributo de cada item e do teste em si, em termos de quantidade de informação, para diferentes níveis de habilidades dos alunos. Com base nestas informações, um importante passo visando a simplificação do processo avaliativo pode ser alcançado: a criação de um banco de itens calibrados. Uma vez calibrados os itens, estes podem ser depositados em um banco a partir do qual o professor tem a prerrogativa de selecionar itens com características pretendidas e construir um teste que lhe proporcione resultados satisfatórios. O professor pode ainda fazer

análise dos itens, removendo os maus, melhorando alguns e adicionando outros. Com um banco de itens calibrados bem equipado e robusto, isto é, abrangendo diversos conteúdos e níveis de complexidade, o professor pode, a partir de uma amostra de itens, estimar o nível de habilidade dos alunos e, facilmente, produzir um teste adequado às diferentes habilidades dos alunos. O conceito de parametrização de exercícios pode ser útil para a não repetição taxativa dos itens, sempre que se desejar usar algum, bem como na criação de itens variantes com as mesmas características.

Uma das grandes limitações deste estudo é o teste em si e o tamanho da amostra. O teste resultou da junção dos 5 momentos de avaliação realizados durante o semestre letivo uma vez que cada um, tinha apenas um máximo de 8 itens e, portanto, as estatísticas resultantes seriam instáveis. O mesmo pode-se dizer em relação à amostra. Inicialmente eram 137 alunos mas somente 111 participaram de todos os momentos de avaliação. Este tamanho de amostra, independentemente do modelo que for a usar para estimar os parâmetros dos itens e as habilidades, está abaixo do recomendado.

Este estudo destacou a meta-avaliação de questões de escolha múltipla com recurso a TCT e a TRI. Impreterivelmente, os professores precisam ter o costume de avaliar os seus instrumentos de avaliação por forma a aperfeiçoá-los cada vez mais e garantir o alcance de objetivos educacionais satisfatórios. Esta recomendação é extensiva a todos os gestores do processo educativo. Para trabalhos futuros, é do nosso interesse fazer meta-avaliação de testes compostos por itens não dicotómicos e abordar os modelos multidimensionais da TRI.

Bibliografia

Altman, D.G. (1991) *Practical Statistics for Medical Research*. New York: Chapman and Hall.

Andrade, D.F., Tavares, H.R., Valle, R.C. (2000) *Teoria da Resposta ao Item: Conceitos e Aplicações*. São Paulo: Associação Brasileira de Estatística - ABE.

Andrich, D.A. (1978) Rating Formulation for Ordered Response Categories. *Psychometrika*, 43(4), 561-73.

Baker, F.B. (1992) *Item Response Theory - Parameter Estimation Techniques*. New York: Marcel Dekker, Inc.

Bock, R.D., Lieberman, M. (1970) Fitting a Response Model for n Dichotomously Scored Items. *Psychometrika*, 35(2), 179-197.

Bock, R.D., Aitkin, M. (1981) Marginal Maximum Likelihood Estimation of Item Parameters: Application of an EM Algorithm. *Psychometrika*, 46(4), 443-459.

Burnham, K.P., Anderson, D.R. (2004) Multimodel Inference: Understanding AIC and BIC in Model Selection. *Sociological Methods and Research*, 33(2), 261-304.

Campbell, D.T., Stanley, J. (1973) *Experimental and Quasi-Experimental Designs for Research*. Skokie, IL: Rand McNally.

Carmines, E.G., Zeller, R.A. (1979). *Reliability and Validity Assessment*. Beverly Hills: CA. Sage University Paper.

Casella, G., Berger, R.L. (2002) *Statistical Inference*. 2nd Edition, Thomson Learning, USA.

Cortina, J. M. (1993) What is Coefficient Alpha? An Examination of Theory and Applications. *Journal of Applied Psychology*, 78, 98-104.

- Costa, P. (2005). Modelos de Resposta ao Item. *Dissertação de Mestrado*. Covilhã: Universidade da Beira Interior.
- Cronbach, L. J. (1951) Coefficient Alpha and the Internal Structure of Tests. *Psychometrika*, 16(3), 297-334.
- Cruz, J.P., Oliveira, M.P., Seabra, D. (2013) Crie o seu Arquivo de Exercícios Patrametrizados. *Gazeta de Matemática*, 170, 26-31.
- Davidson, E.J. (2005) *Evaluation Methodology Basics*. Thousands Oaks, CA: Sage.
- De Araújo, E.A.C., Andrade, D.F., Bortolotti, S.L.V. (2009) Teoria da Resposta ao Item. *Revista da Escola de Enfermagem USP*, São Paulo, 43 (Esp), 1000–1008.
- De Vellis, R. F. (2012). *Scale Development: Theory and Applications*. 3rd Edition, Thousand Oaks, California. Sage Publications.
- Dempster, A.P., Laird, N.M., Rubin, D.B. (1977) Maximum Likelihood from incomplete Data via the EM Algorithm. *Journal of the Royal Statistical Society, Serie B*, 39, 1-38.
- Díaz, C., Batanero, C., Cobo, B. (2003) Fiabilidad y Generalizabilidad. Aplicaciones en Evaluación Educativa. *Números*, 54, 3-21.
- Ebel, R.L. (1979). *Essentials of Educational Measurement*. 3rd ed. Englewood Cliffs, NJ, Prentice-Hall.
- Embreston, S. (1996) The New Rules of Measurement. *Psychological Assessment*, 8(4), 341-349.
- Embretson, S., Reise, S.P. (2000) *Item Response Theory for Psychologists*. New Jersey: Lawrence Erlbaum Associates.
- Fan, X. (1998). Item Response Theory and Classical Test Theory: An Empirical Comparison of their Item/Person Statistics. *Educational and Psychological Measurement*, 58(3), 357-381.
- Fleiss, J. (1981). *Statistical Methods for Rates and Proportions*. 2nd Edition, New York, John Wiley & Sons.
- Fonseca, J. (2001) *Estatística Matemática*. Vol.1, Edições Sílabo, 1^a Edição, Lisboa.

- Fonseca, R., Silva, P., Silva, R. (2007) Acordo Inter-juízes: o caso do coeficiente de kappa. *Laboratório de Psicologia*, 5(1), 81 -90.
- Hambleton, R.K., van der Linden, W.J. (1982). Advances in Item Response Theory and Applications: An Introduction. *Applied Psychological Measurement*, 6(4), 373-378.
- Hambleton, R.K., Jones, R.W. (1993) Comparison of Classical Test Theory and Item Response Theory and their Applications to Test Development. *Educational Measurement: Issues and Practice*, 12(3), 38-47.
- Hambleton, R.K. (1982). *Item response Theory: The Three-Parameter Logistic Model*.
- Hambleton, R.K., Swaminathan, H., Rogers, H.J. (1991). *Fundamentals of Item Response Theory*. North Caroline: Sage Publications.
- Hayes, B.E. (1998) *Measuring Customer Satisfaction: Survey Design, Use, and Statistical Analysis Methods*. Milwaukee, Wisconsin: ASQC Quality Press.
- Hogan, T.P., Agnello, J. (2004) An Empirical Study of Reporting Practices Concerning Measurement Validity. *Educational and Psychological Measurement*, 64(4), 802-812.
- Lord, F. (1980) *Applications of Item Response Theory to Practical Testing Problems*. Hillsdale: Erlbaum. Lord, F.M., Novick, M.R. (1968) *Statistical Theories of Mental Test Scores*. New York: Addison-Wesley.
- Maroco, J., Garcia-Marques, T. (2006) Qual a Fiabilidade do Alfa de Cronbach? Questões Antigas e Soluções Modernas? *Laboratório de Psicologia*, 4(1), 65-90.
- Masters G.N. (1982) A Rasch Model for Partial Credit Scoring. *Psychometrika*, 47(1), 149-74.
- MEGUA (2010) MEGUA – Mathematics Exercise Generator. Code and documentation can be found at <http://code.google.com/p/megua>.
- Muñiz, J. (2010) Las Teorías de los Tests: Teoría Clásica y Teoría de Respuesta a los Ítems. *Papeles del Psicólogo*, 31(1), 57-66.
- Muraki E.A. (1992) Generalized Partial Credit Model: Application of an EM Algorithm. *Applied Psychological Measurement*, 16(1), 159-76.

- Murteira, B., Ribeiro, C.S., e Silva, J.A., Pimenta, C., Pimenta, F. (2015) *Introdução à Estatística*. 3^aed, Escolar Editora, Lisboa.
- Nelder, J.A., Wedderburn, R.W.M. (1972) Generalized Linear Models. *Journal of the Royal Statistical Society. Series A (General)*, 135 (3), 370-384.
- Nocedal, J., Wright, S.J. (2006) *Numerical Optimization*. 2nd Edition, Springer.
- Oliveira, P., Fonseca, A., Ramos, A., Peixoto, E., Gomes, S. (2014) PmatE – 25 Years at the Forefront of Education, *Proceedings 11th International Conference on Hands-On Science*.
- Pasquali, L. (2009) Psicometria. *Revista da Escola de Enfermagem USP*, 43(Esp), 992-9.
- Peterson, R. A. (1994) A Meta-analysis of Cronbach's Coefficient Alpha. *Journal of Consumer Research*, 21(2), 381-391.
- PmatE (2010) An Old Project in Education, Teaching and Learning Using New Technologies. *Proceedings ICERI International Conference of Education , Research and Innovation*, 7249-7253.
- Primi, R. (2012) Psicometria: Fundamentos Matemáticos da Teoria Clássica de Testes. *Avaliação Psicológica*, 11(2), 297-307.
- Rao, S.S. (2009) *Engineering Optimization, Theory and Practice*. 4th Edition, New Jersey, John Wiley & Sons.
- Rasch, G. (1960) *Probabilistic Models for Some Intelligence and Attainment Tests*. Copenhagen: Danish Institute for Educational Research and St. Paul.
- R Core Team (2014). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. <http://www.R-project.org/>.
- Reise, S.P., Ainsworth, A.T., Haviland, M.G. (2005) Item Response Theory: Fundamentals, Applications, and Promise in Psychological Research. *American Psychological Society*, 14(2), 95-101.
- Rizopoulos, D. (2006) ltm: An r Package for Latent Variable Modelling and Item Response Theory Analyses. *Journal of Statistical Software*, 17(5), 1-25.

- Stein, W.A. et al.(2011) Sage Mathematics Software (Version 4.6.1), The Sage Development Team. <http://www.sagemath.org>.
- Scriven, M. (1969) An Introduction to Meta-Evaluation. *Educational Product Report*, 2(5), 36-38.
- Scriven, M. (1991) *Evaluation Thesaurus*. 4th. Edition, Newbury Park, CA: Sage.
- Severo, M., Tavares, M.A.F., (2010) Meta-Evaluation in Clinical Anatomy: A Practical Application. *Anatomical Sciences Education*, 3, 17-24.
- Siegel, S. (1975) *Estatística Não-Paramétrica para Ciências do Comportamento*. McGraw-Hill, São Paulo.
- Stufflebeam, D.L. (2011) Meta-Evaluation. *Journal of MultiDisciplinary Evaluation*, 7(15), 99-158.
- Tavakol, M., Dennick, R. (2011). Post-examination analysis of objective tests. *Medical teacher*, 33(6),447-458.
- Ureña, J., Romera, E.M., Casas, J.A., Viejo, C. (2015) Psychometrics properties of Psychological Dating Violence Questionnaire: A study with young couples. *International Journal of Clinical and Health Psychology*, 15, 52-60.
- van der Linden, W.J., Hambleton, R.K. (1997) *Handbook of Modern Item Response Theory*. New York: Springer-Verlag.

Apêndices

Apêndice A – Tabelas

Tabela A.1: Índices de dificuldade, discriminação e de consistência interna de cada item obtidos via TCT

Item	Nº.Acertos	% Acertos	Corr. P.Bisserial	α (excluindo item i)
1	99	89.19	0.0820	0.8009
2	32	28.83	-0.0147	0.8074
3	106	95.50	0.1318	0.7987
4	76	68.47	0.2291	0.7989
5	70	63.06	0.4134	0.7917
6	83	74.77	0.4095	0.7918
7	39	35.14	0.3039	0.7962
8	55	49.55	0.4828	0.7887
9	77	69.37	0.5676	0.7853
10	95	85.59	0.2447	0.7970
11	53	47.75	0.2392	0.7991
12	71	63.96	0.5747	0.7847
13	47	42.34	0.3614	0.7940
14	49	44.14	0.5118	0.7874
15	46	41.44	0.1917	0.8009
16	56	50.45	0.2247	0.7997
17	57	51.35	0.3776	0.7933
18	66	59.46	0.6586	0.7807

19	71	63.96	0.3430	0.7946
20	61	54.95	0.4316	0.7909
21	43	38.74	0.1863	0.8010
22	89	80.18	0.4568	0.7903
23	59	53.15	0.5513	0.7856
24	98	88.29	0.4229	0.7921
25	83	74.77	0.2388	0.7981
26	89	80.18	0.3173	0.7950
27	100	90.09	0.2065	0.7977
28	72	64.86	0.4533	0.7900
29	95	85.59	0.4426	0.7912
30	66	59.46	0.2307	0.7993
31	70	63.06	0.0357	0.8066
32	65	58.56	0.5138	0.7873
33	62	55.86	0.4339	0.7908
34	64	57.66	0.3358	0.7950
35	74	66.67	0.3080	0.7959
36	97	87.39	0.3981	0.7926
37	78	70.27	0.4354	0.7908

Tabela A.2: Parâmetros de dificuldade, discriminação e acerto casual de cada item estimados via TRI pelos modelos 1PL, 2PL e 3PL

Item	Modelo 1PL	Modelo 2PL		Modelo 3PL		
	Dificuld.	Dificuld.	Discrimin.	Dificuld.	Discrimin.	Casual
1	-2.3750	-26.0867	0.0810	-8.4055	0.2496	0.0135
2	1.0321	-4.4524	-0.2047	-2.2391	-12.8908	0.2761
3	-3.3705	-5.4712	0.5861	-5.7449	0.5117	0.1662
4	-0.8966	-2.7521	0.2866	1.0402	0.8154	0.5239
5	-0.6216	-0.7124	0.8710	-0.6540	0.8082	0.0000
6	-1.2490	-1.2778	1.0277	-0.1259	1.9168	0.4053
7	0.7007	1.0819	0.6204	1.1071	0.6662	0.0000
8	0.0156	0.0259	1.0092	0.1303	1.0773	0.0000
9	-0.9445	-0.6841	1.9328	-0.0004	67.6659	0.3004
10	-2.0189	-3.5069	0.5356	0.7163	26.6218	0.8123
11	0.0989	0.2900	0.3236	0.3395	0.3745	0.0000
12	-0.6661	-0.5026	1.7531	-0.3492	1.9108	0.0000
13	0.3510	0.5030	0.6870	1.2588	10.1713	0.3341
14	0.2664	0.2294	1.4724	0.5144	2.4061	0.1089
15	0.3936	1.4123	0.2494	1.2060	0.3181	0.0002
16	-0.0262	-0.0498	0.3462	1.3670	12.6976	0.4497
17	-0.0677	-0.0661	0.9109	0.0377	0.9354	0.0000
18	-0.4471	-0.2818	3.0272	0.0019	100.2278	0.1064
19	-0.6662	-0.9986	0.6236	0.8678	108.5430	0.5668
20	-0.2349	-0.2519	0.9232	-0.1472	0.9278	0.0000
21	0.5234	2.2558	0.2058	1.6556	59.4857	0.3641
22	-1.5968	-1.2985	1.4887	-0.4956	2.5437	0.3388
23	-0.1512	-0.1138	1.6075	0.0109	1.6550	0.0000
24	-2.2779	-1.7247	1.6740	-0.8998	2.3656	0.4337
25	-1.2488	-2.9695	0.3769	-3.1422	0.3450	0.0004
26	-1.5968	-2.0280	0.7700	0.6709	40.8911	0.7143
27	-2.4791	-5.1989	0.4389	0.6995	1.9927	0.8528

28	-0.7112	-0.7263	1.0260	-0.6242	1.0174	0.0000
29	-2.0189	-1.6011	1.5433	-1.5918	1.4167	0.0000
30	-0.4471	-1.0838	0.3634	0.3281	0.4302	0.2261
31	-0.6217	2.2600	-0.2407	2.5496	-0.2205	0.0000
32	-0.4043	-0.3404	1.3824	0.1389	2.2500	0.1754
33	-0.2770	-0.2864	0.0000	-0.1845	0.9636	0.9794
34	-0.3617	-0.5345	0.6260	0.8974	2.0857	0.4206
35	-0.8028	-1.1805	0.6396	-1.0802	0.6437	0.0000
36	-2.1866	-1.8621	1.3529	-1.7156	1.4419	0.0000
37	-0.9931	-1.0541	0.9738	0.2655	2.0215	0.4525

Tabela A.3: Habilidade de cada aluno estimada via TRI pelos modelos 1PL, 2PL e 3PL

Indivíduo	Modelo 1PL	Modelo 2PL	Modelo 3PL
j1	-0.556	0.022	0.072
j2	0.161	0.478	0.581
j3	0.626	1.049	0.899
j4	0.161	0.580	0.784
j5	-0.275	-0.433	-0.077
j6	0.466	0.542	0.398
j7	-1.112	-1.022	-0.685
j8	-1.832	-1.630	-2.424
j9	-2.297	-2.073	-2.462
j10	-0.275	-0.244	-0.012
j11	1.333	1.075	1.027
j12	0.963	0.496	0.324
j13	-0.973	-1.093	-1.237
j14	-0.275	-0.106	0.047
j15	-1.394	-1.374	-1.234
j16	-0.556	-0.239	0.042
j17	0.013	-0.154	-0.044
j18	-0.132	-0.248	0.039
j19	-1.683	-1.813	-1.452
j20	0.312	0.122	0.060
j21	-0.275	-0.514	-0.069
j22	-0.275	0.187	0.005
j23	-0.416	-0.281	0.003
j24	0.791	0.783	1.690
j25	-0.695	-0.616	-0.230
j26	-1.253	-1.198	-0.747
j27	0.161	-0.368	-0.065
j28	0.791	0.585	0.376
j29	-1.538	-1.242	-0.887

j30	-1.394	-1.486	-1.673
j31	-0.556	-0.280	-0.011
j32	0.791	1.070	1.691
j33	0.626	0.778	0.821
j34	0.791	0.721	0.906
j35	-0.132	0.345	0.370
j36	0.466	1.052	0.675
j37	-0.132	0.100	0.073
j38	1.333	1.614	1.281
j39	-2.138	-1.949	-1.947
j40	-1.112	-1.271	-1.221
j41	-1.112	-1.432	-1.423
j42	-1.538	-1.716	-1.431
j43	0.161	0.363	0.895
j44	0.626	0.835	1.379
j45	-0.416	-0.394	-0.013
j46	-0.556	-0.493	-0.012
j47	-0.416	-0.161	0.059
j48	0.013	-0.295	-0.012
j49	-0.695	-0.788	-0.439
j50	-0.695	-0.881	-0.592
j51	0.312	0.294	0.213
j52	1.534	1.631	1.725
j53	-0.834	-1.195	-1.206
j54	0.626	0.514	0.534
j55	-0.556	-0.495	0.038
j56	-0.834	-0.960	-0.886
j57	-0.275	-0.352	-0.013
j58	0.013	-0.339	-0.013
j59	0.161	-0.337	-0.063
j60	-0.416	-0.297	-0.011

j61	0.312	0.593	0.797
j62	-0.695	-0.877	-0.651
j63	-0.695	-0.341	0.003
j64	-0.695	-0.809	-0.625
j65	-0.275	-0.277	0.003
j66	-0.695	-0.437	-0.012
j67	0.466	0.302	0.401
j68	0.013	0.089	0.190
j69	0.312	0.426	0.893
j70	0.963	0.813	0.904
j71	0.791	0.289	-0.010
j72	1.143	1.210	1.450
j73	0.466	0.206	0.157
j74	1.143	1.350	1.125
j75	0.963	1.241	1.219
j76	0.963	1.238	1.471
j77	0.161	0.360	0.445
j78	0.312	0.310	0.240
j79	0.791	0.804	0.609
j80	0.312	0.594	0.898
j81	0.791	1.249	1.074
j82	0.963	0.978	1.424
j83	1.534	1.562	1.572
j84	1.748	1.644	1.724
j85	1.333	1.547	1.719
j86	-1.253	-1.429	-1.102
j87	0.963	0.879	1.219
j88	-0.834	-0.589	0.033
j89	-0.973	-1.164	-1.488
j90	-0.834	-0.787	-0.619
j91	0.161	0.344	0.563

j92	-0.834	-0.892	-0.558
j93	-1.112	-1.304	-1.179
j94	-0.275	0.050	0.156
j95	0.161	0.134	0.292
j96	-0.132	-0.234	0.003
j97	0.312	-0.009	0.065
j98	0.312	-0.169	0.003
j99	0.013	-0.217	0.003
j100	-0.973	-0.979	0.000
j101	0.161	-0.160	-0.043
j102	0.626	0.763	0.829
j103	0.161	0.001	-0.010
j104	1.748	1.270	1.526
j105	-0.132	-0.519	-0.396
j106	0.312	-0.242	-0.012
j107	-0.275	-0.717	-0.400
j108	1.143	1.093	1.219
j109	1.143	0.508	0.576
j110	1.143	0.809	1.432
j111	1.748	1.656	1.730

Apêndice B – Questionário aplicado aos docentes

QUESTIONÁRIO A DOCENTES

Solicitamos a sua melhor colaboração no preenchimento deste questionário que se enquadra numa investigação no âmbito da minha dissertação para obtenção do grau de mestre em Matemática e Aplicações, Especialização em Estatística e Investigação Operacional, da Universidade de Aveiro. Pretende-se com este questionário colher a avaliação de docentes do Ensino Superior a questões de escolha múltiplas inseridas em provas de avaliação na unidade curricular de Bioestatística realizadas no presente ano letivo. A sua importante colaboração irá contribuir para clarificar se há discrepância na perceção do grau de dificuldade e de discriminação dessas questões e com isso perceber de que modo os docentes podem ir ao encontro das expectativas dos estudantes após o estudo.

As respostas ao presente questionário serão tratadas de forma agregada de modo a garantir o anonimato dos docentes intervenientes. Comprometemos a fazer uso da informação recolhida, exclusivamente, para fins de investigação e eventual publicação de resultados.

Desde já agradecemos pelo seu precioso tempo concedido e pela sua colaboração.

I. DADOS PROFISSIONAIS (SELECIONE A SUA SITUAÇÃO)

1.1 Habilitações

☐ Mestrado

☐ Doutoramento

☐ Com Agregação

☐ Convidado

☐ Contratado

☐ Efetivo

1.2 Vínculo profissional

☐ Outro

1.3 Tempo de serviço (anos)

☐ Até 5

☐ De 6 a 10

☐ De 11 a 20

☐ Acima de 20

II. AVALIAÇÃO POR QUESTÃO

Para cada uma das questões da prova em anexo assinale, na escala de 1 a 5, o nível de concordância que melhor reflecte a característica pedida, tendo por base aquilo que, na sua opinião, seria sensato para uma disciplina de Bioestatística para alunos do 1º ano da licenciatura de Biologia, e tendo em conta a seguinte escala: 1 – Discordo totalmente; 2 – Discordo parcialmente; 3 – Não concordo nem discordo; 4 – Concordo parcialmente; 5 – Concordo. Assinale NS sempre que não souber

Questão 1						
Clareza de linguagem do enunciado e das 4 opções (globalmente)	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/> 5	<input type="checkbox"/> NS
Grau de dificuldade da questão	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/> 5	<input type="checkbox"/> NS
Adequação da questão (ao conteúdo curricular expectável)	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/> 5	<input type="checkbox"/> NS
Capacidade de discriminar (alunos muito bons, que estudam, dos restantes).	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/> 5	<input type="checkbox"/> NS

Questão 2						
Clareza de linguagem do enunciado e das 4 opções (globalmente)	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/> 5	<input type="checkbox"/> NS
Grau de dificuldade da questão	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/> 5	<input type="checkbox"/> NS
Adequação da questão (ao conteúdo curricular expectável)	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/> 5	<input type="checkbox"/> NS
Capacidade de discriminar (alunos muito bons, que estudam, dos restantes).	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/> 5	<input type="checkbox"/> NS

Questão 3						
Clareza de linguagem do enunciado e das 4 opções (globalmente)	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/> 5	<input type="checkbox"/> NS
Grau de dificuldade da questão	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/> 5	<input type="checkbox"/> NS
Adequação da questão (ao conteúdo curricular expectável)	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/> 5	<input type="checkbox"/> NS
Capacidade de discriminar (alunos muito bons, que estudam, dos restantes).	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/> 5	<input type="checkbox"/> NS

Questão 4						
Clareza de linguagem do enunciado e das 4 opções (globalmente)	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/> 5	<input type="checkbox"/> NS
Grau de dificuldade da questão	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/> 5	<input type="checkbox"/> NS
Adequação da questão (ao conteúdo curricular expectável)	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/> 5	<input type="checkbox"/> NS
Capacidade de discriminar (alunos muito bons, que estudam, dos restantes).	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/> 5	<input type="checkbox"/> NS

Questão 5						
Clareza de linguagem do enunciado e das 4 opções (globalmente)	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/> 5	<input type="checkbox"/> NS
Grau de dificuldade da questão	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/> 5	<input type="checkbox"/> NS
Adequação da questão (ao conteúdo curricular expectável)	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/> 5	<input type="checkbox"/> NS
Capacidade de discriminar (alunos muito bons, que estudam, dos restantes).	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/> 5	<input type="checkbox"/> NS

Questão 6						
Clareza de linguagem do enunciado e das 4 opções (globalmente)	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/> 5	<input type="checkbox"/> NS
Grau de dificuldade da questão	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/> 5	<input type="checkbox"/> NS
Adequação da questão (ao conteúdo curricular expectável)	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/> 5	<input type="checkbox"/> NS
Capacidade de discriminar (alunos muito bons, que estudam, dos restantes).	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/> 5	<input type="checkbox"/> NS

Questão 7						
Clareza de linguagem do enunciado e das 4 opções (globalmente)	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/> 5	<input type="checkbox"/> NS
Grau de dificuldade da questão	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/> 5	<input type="checkbox"/> NS
Adequação da questão (ao conteúdo curricular expectável)	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/> 5	<input type="checkbox"/> NS
Capacidade de discriminar (alunos muito bons, que estudam, dos restantes).	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/> 5	<input type="checkbox"/> NS

Questão 8						
Clareza de linguagem do enunciado e das 4 opções (globalmente)	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/> 5	<input type="checkbox"/> NS
Grau de dificuldade da questão	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/> 5	<input type="checkbox"/> NS
Adequação da questão (ao conteúdo curricular expectável)	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/> 5	<input type="checkbox"/> NS
Capacidade de discriminar (alunos muito bons, que estudam, dos restantes).	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/> 5	<input type="checkbox"/> NS

Questão 9						
Clareza de linguagem do enunciado e das 4 opções (globalmente)	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/> 5	<input type="checkbox"/> NS
Grau de dificuldade da questão	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/> 5	<input type="checkbox"/> NS
Adequação da questão (ao conteúdo curricular expectável)	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/> 5	<input type="checkbox"/> NS
Capacidade de discriminar (alunos muito bons, que estudam, dos restantes).	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/> 5	<input type="checkbox"/> NS

Questão 10						
Clareza de linguagem do enunciado e das 4 opções (globalmente)	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/> 5	<input type="checkbox"/> NS
Grau de dificuldade da questão	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/> 5	<input type="checkbox"/> NS
Adequação da questão (ao conteúdo curricular expectável)	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/> 5	<input type="checkbox"/> NS
Capacidade de discriminar (alunos muito bons, que estudam, dos restantes).	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/> 5	<input type="checkbox"/> NS

Questão 11

Clareza de linguagem do enunciado e das 4 opções (globalmente)	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/> 5	<input type="checkbox"/> NS
Grau de dificuldade da questão	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/> 5	<input type="checkbox"/> NS
Adequação da questão (ao conteúdo curricular expectável)	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/> 5	<input type="checkbox"/> NS
Capacidade de discriminar (alunos muito bons, que estudam, dos restantes).	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/> 5	<input type="checkbox"/> NS

Questão 12

Clareza de linguagem do enunciado e das 4 opções (globalmente)	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/> 5	<input type="checkbox"/> NS
Grau de dificuldade da questão	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/> 5	<input type="checkbox"/> NS
Adequação da questão (ao conteúdo curricular expectável)	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/> 5	<input type="checkbox"/> NS
Capacidade de discriminar (alunos muito bons, que estudam, dos restantes).	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/> 5	<input type="checkbox"/> NS

Questão 13

Clareza de linguagem do enunciado e das 4 opções (globalmente)	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/> 5	<input type="checkbox"/> NS
Grau de dificuldade da questão	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/> 5	<input type="checkbox"/> NS
Adequação da questão (ao conteúdo curricular expectável)	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/> 5	<input type="checkbox"/> NS
Capacidade de discriminar (alunos muito bons, que estudam, dos restantes).	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/> 5	<input type="checkbox"/> NS

Questão 14

Clareza de linguagem do enunciado e das 4 opções (globalmente)	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/> 5	<input type="checkbox"/> NS
Grau de dificuldade da questão	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/> 5	<input type="checkbox"/> NS
Adequação da questão (ao conteúdo curricular expectável)	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/> 5	<input type="checkbox"/> NS
Capacidade de discriminar (alunos muito bons, que estudam, dos restantes).	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/> 5	<input type="checkbox"/> NS

Questão 15

Clareza de linguagem do enunciado e das 4 opções (globalmente)	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/> 5	<input type="checkbox"/> NS
Grau de dificuldade da questão	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/> 5	<input type="checkbox"/> NS
Adequação da questão (ao conteúdo curricular expectável)	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/> 5	<input type="checkbox"/> NS
Capacidade de discriminar (alunos muito bons, que estudam, dos restantes).	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/> 5	<input type="checkbox"/> NS

Questão 16

Clareza de linguagem do enunciado e das 4 opções (globalmente)	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/> 5	<input type="checkbox"/> NS
Grau de dificuldade da questão	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/> 5	<input type="checkbox"/> NS
Adequação da questão (ao conteúdo curricular expectável)	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/> 5	<input type="checkbox"/> NS
Capacidade de discriminar (alunos muito bons, que estudam, dos restantes).	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/> 5	<input type="checkbox"/> NS

Questão 17

Clareza de linguagem do enunciado e das 4 opções (globalmente)	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/> 5	<input type="checkbox"/> NS
Grau de dificuldade da questão	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/> 5	<input type="checkbox"/> NS
Adequação da questão (ao conteúdo curricular expectável)	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/> 5	<input type="checkbox"/> NS
Capacidade de discriminar (alunos muito bons, que estudam, dos restantes).	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/> 5	<input type="checkbox"/> NS

Questão 18

Clareza de linguagem do enunciado e das 4 opções (globalmente)	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/> 5	<input type="checkbox"/> NS
Grau de dificuldade da questão	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/> 5	<input type="checkbox"/> NS
Adequação da questão (ao conteúdo curricular expectável)	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/> 5	<input type="checkbox"/> NS
Capacidade de discriminar (alunos muito bons, que estudam, dos restantes).	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/> 5	<input type="checkbox"/> NS

Questão 19

Clareza de linguagem do enunciado e das 4 opções (globalmente)	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/> 5	<input type="checkbox"/> NS
Grau de dificuldade da questão	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/> 5	<input type="checkbox"/> NS
Adequação da questão (ao conteúdo curricular expectável)	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/> 5	<input type="checkbox"/> NS
Capacidade de discriminar (alunos muito bons, que estudam, dos restantes).	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/> 5	<input type="checkbox"/> NS

Questão 20

Clareza de linguagem do enunciado e das 4 opções (globalmente)	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/> 5	<input type="checkbox"/> NS
Grau de dificuldade da questão	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/> 5	<input type="checkbox"/> NS
Adequação da questão (ao conteúdo curricular expectável)	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/> 5	<input type="checkbox"/> NS
Capacidade de discriminar (alunos muito bons, que estudam, dos restantes).	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/> 5	<input type="checkbox"/> NS

Questão 21

Clareza de linguagem do enunciado e das 4 opções (globalmente)	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/> 5	<input type="checkbox"/> NS
Grau de dificuldade da questão	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/> 5	<input type="checkbox"/> NS
Adequação da questão (ao conteúdo curricular expectável)	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/> 5	<input type="checkbox"/> NS
Capacidade de discriminar (alunos muito bons, que estudam, dos restantes).	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/> 5	<input type="checkbox"/> NS

Questão 22

Clareza de linguagem do enunciado e das 4 opções (globalmente)	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/> 5	<input type="checkbox"/> NS
Grau de dificuldade da questão	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/> 5	<input type="checkbox"/> NS
Adequação da questão (ao conteúdo curricular expectável)	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/> 5	<input type="checkbox"/> NS
Capacidade de discriminar (alunos muito bons, que estudam, dos restantes).	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/> 5	<input type="checkbox"/> NS

Questão 23

Clareza de linguagem do enunciado e das 4 opções (globalmente)	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/> 5	<input type="checkbox"/> NS
Grau de dificuldade da questão	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/> 5	<input type="checkbox"/> NS
Adequação da questão (ao conteúdo curricular expectável)	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/> 5	<input type="checkbox"/> NS
Capacidade de discriminar (alunos muito bons, que estudam, dos restantes).	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/> 5	<input type="checkbox"/> NS

Questão 24

Clareza de linguagem do enunciado e das 4 opções (globalmente)	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/> 5	<input type="checkbox"/> NS
Grau de dificuldade da questão	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/> 5	<input type="checkbox"/> NS
Adequação da questão (ao conteúdo curricular expectável)	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/> 5	<input type="checkbox"/> NS
Capacidade de discriminar (alunos muito bons, que estudam, dos restantes).	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/> 5	<input type="checkbox"/> NS

Questão 25

Clareza de linguagem do enunciado e das 4 opções (globalmente)	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/> 5	<input type="checkbox"/> NS
Grau de dificuldade da questão	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/> 5	<input type="checkbox"/> NS
Adequação da questão (ao conteúdo curricular expectável)	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/> 5	<input type="checkbox"/> NS
Capacidade de discriminar (alunos muito bons, que estudam, dos restantes).	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/> 5	<input type="checkbox"/> NS

Questão 26

Clareza de linguagem do enunciado e das 4 opções (globalmente)	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/> 5	<input type="checkbox"/> NS
Grau de dificuldade da questão	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/> 5	<input type="checkbox"/> NS
Adequação da questão (ao conteúdo curricular expectável)	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/> 5	<input type="checkbox"/> NS
Capacidade de discriminar (alunos muito bons, que estudam, dos restantes).	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/> 5	<input type="checkbox"/> NS

Questão 27

Clareza de linguagem do enunciado e das 4 opções (globalmente)	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/> 5	<input type="checkbox"/> NS
Grau de dificuldade da questão	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/> 5	<input type="checkbox"/> NS
Adequação da questão (ao conteúdo curricular expectável)	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/> 5	<input type="checkbox"/> NS
Capacidade de discriminar (alunos muito bons, que estudam, dos restantes).	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/> 5	<input type="checkbox"/> NS

Questão 28

Clareza de linguagem do enunciado e das 4 opções (globalmente)	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/> 5	<input type="checkbox"/> NS
Grau de dificuldade da questão	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/> 5	<input type="checkbox"/> NS
Adequação da questão (ao conteúdo curricular expectável)	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/> 5	<input type="checkbox"/> NS
Capacidade de discriminar (alunos muito bons, que estudam, dos restantes).	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/> 5	<input type="checkbox"/> NS

Questão 29

Clareza de linguagem do enunciado e das 4 opções (globalmente)	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/> 5	<input type="checkbox"/> NS
Grau de dificuldade da questão	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/> 5	<input type="checkbox"/> NS
Adequação da questão (ao conteúdo curricular expectável)	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/> 5	<input type="checkbox"/> NS
Capacidade de discriminar (alunos muito bons, que estudam, dos restantes).	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/> 5	<input type="checkbox"/> NS

Questão 30

Clareza de linguagem do enunciado e das 4 opções (globalmente)	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/> 5	<input type="checkbox"/> NS
Grau de dificuldade da questão	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/> 5	<input type="checkbox"/> NS
Adequação da questão (ao conteúdo curricular expectável)	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/> 5	<input type="checkbox"/> NS
Capacidade de discriminar (alunos muito bons, que estudam, dos restantes).	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/> 5	<input type="checkbox"/> NS

Questão 31

Clareza de linguagem do enunciado e das 4 opções (globalmente)	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/> 5	<input type="checkbox"/> NS
Grau de dificuldade da questão	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/> 5	<input type="checkbox"/> NS
Adequação da questão (ao conteúdo curricular expectável)	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/> 5	<input type="checkbox"/> NS
Capacidade de discriminar (alunos muito bons, que estudam, dos restantes).	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/> 5	<input type="checkbox"/> NS

Questão 32

Clareza de linguagem do enunciado e das 4 opções (globalmente)	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/> 5	<input type="checkbox"/> NS
Grau de dificuldade da questão	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/> 5	<input type="checkbox"/> NS
Adequação da questão (ao conteúdo curricular expectável)	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/> 5	<input type="checkbox"/> NS
Capacidade de discriminar (alunos muito bons, que estudam, dos restantes).	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/> 5	<input type="checkbox"/> NS

Questão 33

Clareza de linguagem do enunciado e das 4 opções (globalmente)	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/> 5	<input type="checkbox"/> NS
Grau de dificuldade da questão	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/> 5	<input type="checkbox"/> NS
Adequação da questão (ao conteúdo curricular expectável)	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/> 5	<input type="checkbox"/> NS
Capacidade de discriminar (alunos muito bons, que estudam, dos restantes).	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/> 5	<input type="checkbox"/> NS

Questão 34

Clareza de linguagem do enunciado e das 4 opções (globalmente)	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/> 5	<input type="checkbox"/> NS
Grau de dificuldade da questão	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/> 5	<input type="checkbox"/> NS
Adequação da questão (ao conteúdo curricular expectável)	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/> 5	<input type="checkbox"/> NS
Capacidade de discriminar (alunos muito bons, que estudam, dos restantes).	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/> 5	<input type="checkbox"/> NS

Questão 35

Clareza de linguagem do enunciado e das 4 opções (globalmente)	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/> 5	<input type="checkbox"/> NS
Grau de dificuldade da questão	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/> 5	<input type="checkbox"/> NS
Adequação da questão (ao conteúdo curricular expectável)	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/> 5	<input type="checkbox"/> NS
Capacidade de discriminar (alunos muito bons, que estudam, dos restantes).	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/> 5	<input type="checkbox"/> NS

Questão 36						
Clareza de linguagem do enunciado e das 4 opções (globalmente)	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/> 5	<input type="checkbox"/> NS
Grau de dificuldade da questão	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/> 5	<input type="checkbox"/> NS
Adequação da questão (ao conteúdo curricular expectável)	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/> 5	<input type="checkbox"/> NS
Capacidade de discriminar (alunos muito bons, que estudam, dos restantes).	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/> 5	<input type="checkbox"/> NS

Questão 37						
Clareza de linguagem do enunciado e das 4 opções (globalmente)	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/> 5	<input type="checkbox"/> NS
Grau de dificuldade da questão	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/> 5	<input type="checkbox"/> NS
Adequação da questão (ao conteúdo curricular expectável)	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/> 5	<input type="checkbox"/> NS
Capacidade de discriminar (alunos muito bons, que estudam, dos restantes).	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/> 5	<input type="checkbox"/> NS

III. AVALIAÇÃO GLOBAL DO TESTE

Considere os momentos de avaliação seguintes:

Momento de Avaliação 1: Estatística descritiva (questões 1 a 5)

Momento de Avaliação 2: Probabilidades e intervalos de confiança (questões 6 a 13)

Momento de Avaliação 3: Intervalos de confiança e teste de hipóteses (questões 14 a 21)

Momento de Avaliação 4: ANOVA (questões 22 a 29)

Momento de Avaliação 5: Regressão linear (questões 30 a 37)

3.1 Nas afirmações que se seguem assinale, na escala de 1 a 5, o nível de concordância que melhor reflete a sua opinião resultante exclusivamente de sua apreciação global e tendo em conta a seguinte escala: 1 – Discordo totalmente; 2 – Discordo parcialmente; 3 – Não concordo nem discordo; 4 – Concordo parcialmente; 5 – Concordo. Assinale NS sempre que não souber.

Momento de Avaliação 1						
As questões apresentam uma linguagem concisa	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/> 5	<input type="checkbox"/> 6
Nesta avaliação existe equilíbrio em termos de haver perguntas fáceis, intermédias e difíceis.	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/> 5	<input type="checkbox"/> 6
Os níveis de exigência curricular das questões avaliadas são apropriados	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/> 5	<input type="checkbox"/> 6
A avaliação não contém questões que permitem discriminar os melhores alunos	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/> 5	<input type="checkbox"/> 6

Momento de Avaliação 2						
As questões apresentam uma linguagem concisa	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/> 5	<input type="checkbox"/> 6
Nesta avaliação existe equilíbrio em termos de haver perguntas fáceis, intermédias e difíceis.	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/> 5	<input type="checkbox"/> 6
Os níveis de exigência curricular das questões avaliadas são apropriados	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/> 5	<input type="checkbox"/> 6
A avaliação não contém questões que permitem discriminar os melhores alunos	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/> 5	<input type="checkbox"/> 6

Momento de Avaliação 3						
As questões apresentam uma linguagem concisa	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/> 5	<input type="checkbox"/> 6
Nesta avaliação existe equilíbrio em termos de haver perguntas fáceis, intermédias e difíceis.	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/> 5	<input type="checkbox"/> 6
Os níveis de exigência curricular das questões avaliadas são apropriados	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/> 5	<input type="checkbox"/> 6
A avaliação não contém questões que permitem discriminar os melhores alunos	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/> 5	<input type="checkbox"/> 6

Momento de Avaliação 4						
As questões apresentam uma linguagem concisa	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/> 5	<input type="checkbox"/> 6
Nesta avaliação existe equilíbrio em termos de haver perguntas fáceis, intermédias e difíceis.	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/> 5	<input type="checkbox"/> 6
Os níveis de exigência curricular das questões avaliadas são apropriados	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/> 5	<input type="checkbox"/> 6
A avaliação não contém questões que permitem discriminar os melhores alunos	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/> 5	<input type="checkbox"/> 6

Momento de Avaliação 5						
As questões apresentam uma linguagem concisa	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/> 5	<input type="checkbox"/> 6
Nesta avaliação existe equilíbrio em termos de haver perguntas fáceis, intermédias e difíceis.	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/> 5	<input type="checkbox"/> 6
Os níveis de exigência curricular das questões avaliadas são apropriados	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/> 5	<input type="checkbox"/> 6
A avaliação não contém questões que permitem discriminar os melhores alunos	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/> 5	<input type="checkbox"/> 6

3.2 Para **cada momento de avaliação** queira, por favor, indicar uma questão que considera mais fácil e outra que considera mais difícil.

Momento de avaliação 1	Questão mais fácil	<input type="text"/>	Questão mais difícil	<input type="text"/>
Momento de avaliação 2	Questão mais fácil	<input type="text"/>	Questão mais difícil	<input type="text"/>
Momento de avaliação 3	Questão mais fácil	<input type="text"/>	Questão mais difícil	<input type="text"/>
Momento de avaliação 4	Questão mais fácil	<input type="text"/>	Questão mais difícil	<input type="text"/>
Momento de avaliação 5	Questão mais fácil	<input type="text"/>	Questão mais difícil	<input type="text"/>

3.3 Para **o teste no global** queira, por favor, indicar uma questão que considera mais fácil e outra que considera mais difícil.

Questão mais fácil	<input type="text"/>	Questão mais difícil	<input type="text"/>
--------------------	----------------------	----------------------	----------------------

Obrigado pela colaboração,

Jeremias Jasse

Anexo I – Programa de Bioestatística



BIOESTATÍSTICA

PROGRAMA 2015/16

1. Introdução à Bioestatística e Análise Exploratória de Dados

- Estatística na Biologia.
- Tipos de dados.
 - i. Tipos de variáveis;
 - ii. Escalas de medida.
- Análise Exploratória de Dados.
 - i. Medição de variáveis;
 - ii. Organização de dados em tabelas e gráficos;
 - iii. Medidas de localização, dispersão e forma.

2. Distribuições de Probabilidade

- Variável aleatória, função de distribuição.
- Variáveis aleatórias discretas e contínuas: função massa de probabilidade e função densidade de probabilidade.
- Parâmetros de uma distribuição: valor esperado, variância, momentos, moda, mediana e quantis.
- Distribuições com especial importância: binomial, Poisson, normal.

3. Estatística Inferencial

- Amostragem: noções gerais de amostragem.
 - i. Distribuições de amostragem;
 - ii. Teorema Limite Central.
- Estimação pontual: estimador e estimativa.
- Estimação intervalar: noções gerais sobre intervalos de confiança.
 - i. Intervalos de confiança em populações normais: médias e variâncias;
 - ii. Intervalo de confiança para a proporção.
- Testes de hipóteses paramétricos.
 - i. Noções gerais sobre teste de hipóteses: tipos de hipóteses, erro de 1ª e de 2ª espécie, potência do teste e valor p;
 - ii. Testes de hipóteses paramétricos em populações normais:
 - para a média,



BIOESTATÍSTICA

PROGRAMA 2015/16

- para a variância,
- para a comparação de médias,
- para a comparação de variâncias.
- iii. Teste para a proporção e teste para a comparação de proporções.
- o Validação de pressupostos.
 - i. QQ-plots;
 - ii. Testes de ajustamento à uma distribuição normal: teste de Kolmogorov-Smirnov com correção de Lilliefors, teste de Shapiro Wilk;
 - iii. Teste para a igualdade de variâncias.

4. Análise de Variância

- o Planeamento de experiências com um factor.
 - i. Modelo de efeitos fixos: Pressupostos do modelo, Interpretação dos resultados, Comparações múltiplas;
 - ii. Planeamento de experiências com dois factores, com dois ou mais níveis cada um (Blocos e Observações repetidas)

5. Correlação e Regressão

- o Análise da correlação.
 - i. Medidas de associação variáveis não nominais;
 - Correlação de Pearson,
 - Correlação de Spearman.
 - ii. Diagramas de dispersão;
 - iii. Transformações de dados;
 - iv. Testes para coeficientes de correlação.
- o Regressão linear.
 - i. Apresentação e interpretação do modelo;
 - ii. Estimação pontual dos parâmetros da recta de regressão;
 - iii. Predição de uma observação futura.



BIOESTATÍSTICA

PROGRAMA 2015/16

- iv. Inferência sobre os parâmetros do modelo: intervalos de confiança e testes de hipótese;
- v. O coeficiente de determinação como indicador da qualidade do ajustamento;
- vi. Validação de pressupostos.

6. Análise de dados qualitativos

- i. Tabelas de contingência;
- ii. Teste do χ^2 para a independência;
- iii. Teste de ajustamento do χ^2 .

Aveiro, 4 de fevereiro de 2016

A responsável pela Unidade Curricular,

Anexo II – Teste aplicado aos alunos



+1/1/60+

Questões de Bioestatística ©

Ano Lectivo 2015 / 2016

Mini Teste 1

Q1 Considere o seguinte conjunto de dados de nível concentração sérica (em g / ml) de Gentamicina no sangue recolhido a partir de uma amostra casual de 9 ovelhas:

concentração sérica (g/ml)	33	23	31	32	34	25	34	27	25
----------------------------	----	----	----	----	----	----	----	----	----

Tem-se que:

- ☐ a moda é 2 pois é a frequência que se observa mais vezes (níveis 25 e 34).
- ☐ a amplitude interquartil é igual a 3.
- ☐ mais de 75% das ovelhas apresentam concentração sérica de Gentamicina superior a 33.5.
- ☐ pelo menos 50% das ovelhas apresentam concentração sérica de Gentamicina no sangue acima da média.

Q2 Analisaram-se 200 amostras de água recolhidas numa certa pateira tendo sido registado a concentração de nutrientes. Dos dados conclui-se que:

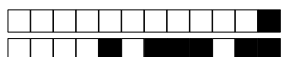
$$\begin{aligned}\text{Percentil de ordem 25} &= 0.4\text{gr}/\text{cm}^3. \\ 3^\circ\text{Quartil} = \text{Média} &= 0.5\text{gr}/\text{cm}^3\end{aligned}$$

Relativamente a este conjunto de dados tem-se que:

- ☐ o comprimento da caixa de bigodes (sem os ditos "bigodes") é igual a 0.5.
- ☐ a se observarem, na caixa de bigodes, níveis atípicos de concentração de nutrientes eles corresponderão a amostras com níveis de concentração superiores a 0.55.
- ☐ a percentagem de observações superiores ou iguais à média é não inferior a 25%.
- ☐ a mediana dos dados será necessariamente um valor superior a 0.4 e inferior a 0.5.

Q3 Considere um estudo para avaliar a relação entre a obesidade e a procriação em ratos da linhagem *wistar*. Nesse estudo analisou-se o número de ratinhos gerados por ninhada (variável 1), o peso da progenitora (variável 2), a idade (em dias) da progenitora (variável 3), o estado de stress (com os níveis baixo, médio e alto) (variável 4), entre outras variáveis.

- ☐ A variável 2 é classificada como quantitativa, contínua, numa escala positiva.
- ☐ A variável 3 é classificada como qualitativa, discreta, numa escala numérica.
- ☐ A variável 4 é classificada como nominal numa escala discreta.
- ☐ A variável 1 é classificada como quantitativa, discreta, numa escala de razão.



+1/2/59+

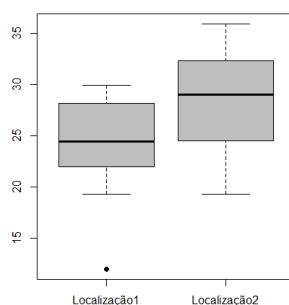
Q4 Num estudo ecológico realizado em rios portugueses mediu-se, durante 45 dias não consecutivos casualmente seleccionados, a concentração de estrôncio (mg/ml) no curso de água do rio Vouga. Os dados foram organizados numa tabela de frequências:

concentração	36	37	38	39	40	41	42	45
Nº de dias	3	11	10	8	9	1	2	1

Relativamente à caixa de bigodes associada a este conjunto de dados:

- ☐ não existem observações atípicas e o "bigode" que se observa do lado direito tem comprimento igual a 5.
- ☐ existe uma observação atípica e o "bigode" que se observa do lado direito tem comprimento igual a 3.
- ☐ existe uma observação atípica e o "bigode" que se observa do lado direito tem comprimento igual a 2.
- ☐ existe uma observação atípica e o "bigode" que se observa do lado direito tem comprimento igual a 4.5.

Q5 É sabido que a qualidade da água não se mantém constante com o tempo e varia de zona para zona. Em duas zonas em estudo (Localização 1 e Localização 2) foram recolhidas várias amostras e medido o índice de salinidade da água em cada amostra. Com base nos valores observados foram obtidas no R as seguintes caixas de bigodes comparativas dos índices de salinidade amostrados nas duas zonas:



Comente as caixas de bigodes, interpretando no contexto do problema o que observa quanto à localização central, dispersão e assimetria da distribuição do índice de salinidade entre as duas zonas;

☐ 0 ☐ 1 ☐ 2 ☐ 3 ☐ 4



+1/3/58+

Mini Teste 2

Q6 A probabilidade de uma ovelha adulta apresentar concentração sérica de Gentamicina elevada no sangue é 0.309. Qual a probabilidade de, numa amostra de 10 ovelhas selecionadas ao acaso, 6 ou mais apresentarem concentração sérica de Gentamicina elevada no sangue?

- ☐ $\approx 6 \times 0.309$
☐ ≈ 0.054
☐ ≈ 0.987
☐ ≈ 0.042

Q7 O nível Y de concentração de um dado composto é bem modelado por uma distribuição de probabilidade $N(\mu = 30, \sigma^2 = 4)$. Então

- ☐ $E(2Y) = 60$ e $Var(2Y) = 8$
☐ $E(2Y) = 30$ e $Var(2Y) = 2$
☐ $E(2Y) = 60$ e $Var(2Y) = 16$
☐ $E(2Y) = 30$ e $Var(2Y) = 4$

Q8 A abundância (X) da espécie *Columbia livia* (pombo-comum) por $10m^2$ numa certa zona da cidade de Aveiro segue uma distribuição de Poisson de parâmetro 4.5. Logo, nessa zona

- ☐ esperam-se encontrar 9 pombos-comum por cada $20m^2$ sendo que $P(X = 0) \approx 0.0111$.
☐ esperam-se encontrar 9 pombos-comum por cada $100m^2$ sendo que $Var(X) = 4.5$.
☐ será pouco provável encontrar mais de 10 pombos-comum por m^2 sendo o valor dessa probabilidade ≈ 0.007 .
☐ com área total de $500m^2$ esperam um total de 500 pombos-comum.

Q9 Num estudo para avaliar características de ratos da linhagem *wistar* sabe-se que o peso X ao desmame de uma cria é bem modelado por uma distribuição normal de média 40g e desvio padrão igual a 5g. A probabilidade de uma ninhada de 5 crias pesar ao desmame mais 220g é

- ☐ ≈ 0.44
☐ ≈ 0.21
☐ ≈ 0.04
☐ ≈ 0.20

Q10 De estudos ecológicos realizados em rios portugueses sabe-se que a concentração de estrôncio (mg/ml), no curso de água do rio Vouga, tem média igual a 40mg/ml e desvio padrão igual a 3.2mg/ml. Nestas condições, em 100 unidades causais de 1ml cada, recolhidas no rio Vouga, a probabilidade da média amostral das 100 unidades variar entre 39.0 e 40.5 é aproximadamente igual a

- ☐ 0.32
☐ 0.77
☐ 0.94
☐ 0.06



+1/4/57+

Q11 Considere as seguintes afirmações:

- A. O 1º quartil de uma distribuição $N(\mu = 3, \sigma^2 = 4)$ é aproximadamente 1.65 .
- B. Uma dada amostra fornece $[10, 20]$ como intervalo a 90% de confiança para μ de uma população normal. Se aumentássemos a confiança para 95%, então o valor do seu limite inferior será superior a 10.

Tem-se

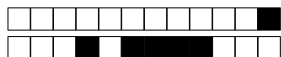
- ☐ A é falsa e B é verdadeira.
- ☐ A é verdadeira e B é falsa.
- ☐ A é verdadeira e B é verdadeira.
- ☐ A é falsa e B é falsa.

Q12 De estudos ecológicos realizados em rios portugueses sabe-se que a concentração de estrôncio (mg/ml), no curso da água do rio, é bem modelado por uma distribuição normal. Com o objetivo de comparar a concentração de estrôncio no rio Vouga e no rio Tejo foram recolhidas 50 amostras de cada rio. Das amostras do rio Vouga registou-se uma média de 38mg/ml com um desvio padrão corrigido de 1.2mg/ml. Da amostra do rio Tejo registou-se uma média de 39.3 mg/ml com um desvio padrão corrigido de 1mg/ml. Um intervalo de confiança a 95% para a diferença de concentrações médias de estrôncio entre os dois rios é:

- ☐ $[-1.873, -0.8672]$
- ☐ $[-1.733, -0.8670]$
- ☐ $[-1.738, -0.8616]$
- ☐ $[-1.739, -0.8614]$

Q13 Um estudo avaliou a obesidade e a procriação de gatos domésticos. Numa amostra de 110 ninhadas de gatos domésticos, verificou-se que 70% das crias nascidas vivas tinham peso abaixo do peso desejado. Assim, uma estimativa intervalar 90% de confiança para a proporção de crias nascidas vivas com peso abaixo do desejado é:

- ☐ $[0.614, 0.786]$.
- ☐ $[0.561, 0.712]$.
- ☐ $[0.628, 0.772]$.
- ☐ $[0.543, 0.857]$.



+1/5/56+

Mini Teste 3

Q14 Os comprimentos das caudas de ratos do campo são bem modelados por uma distribuição normal de variância $\sigma^2 = 0.25$ (parâmetro populacional). Quantos ratos do campos devem ser selecionado para garantir que o intervalo, a 95% confiança para comprimento médio da cauda, construído à custa dessa amostra tem amplitude igual a 0.5?

- ☐ 24
☐ 15
☐ 5
☐ 40

Q15 De estudos ecológicos realizados em rios portugueses sabe-se que a concentração de estrôncio (mg/ml) no curso da água do rio Tejo é bem modelada uma distribuição normal de variância desconhecida σ^2 . Foram recolhidas 51 amostras do rio Tejo tendo-se registado uma média de 39.3 mg/ml de concentração de estrôncio com um desvio padrão corrigido de 1.2mg/ml. Tendo em conta os seguintes resultados obtidos do R e com base na amostra, um intervalo a 95% de confiança para σ^2 é:

```
> qchisq(0.95,50)      > qchisq(0.05,50)      > qchisq(0.95,49)      > qchisq(0.05,49)
[1] 67.50481            [1] 34.76425            [1] 66.33865            [1] 33.93031
> qchisq(0.975,50)     > qchisq(0.025,50)     > qchisq(0.975,49)     > qchisq(0.025,49)
[1] 71.4202             [1] 32.35736            [1] 70.22241            [1] 31.55492
```

- ☐ [1.025, 2.282]
☐ [0.988, 2.180]
☐ [1.008, 2.225]
☐ [1.005, 2.236]

Q16 Com o objetivo de investigar a diversidade e abundância de aves em ambiente urbano, foram realizados censos de avifauna por pontos durante a época de nidificação em várias zonas da cidades de Aveiro. Numa dessas zonas, das 560 aves omnívoras observadas, contabilizaram-se 302 da espécie *Columbia livia* (pombo-comum). Os dados permitem então concluir, ao nível de significância 10%, que a percentagem de pombos-comuns naquela zona de Aveiro é significativamente superior a 50% já que a amostra conduziu ao valor p, do teste de hipótese em causa,

- ☐ ≈ 0.031
☐ ≈ 0.539
☐ ≈ 0.000
☐ ≈ 0.063

Q17 Num estudo pretende-se estudar características de ratos da linhagem *wistar*. Uma dessas características corresponde ao seu peso ao desmame, a qual se assume seguir uma distribuição normal. A experiência contou com 50 ninhadas, com pelo menos 3 crias, observadas em laboratório. De cada ninhada selecionou-se uma cria ao acaso e registou-se o seu peso ao desmame. Com base na amostra dos 50 pesos obteve-se o intervalo [36, 43] como intervalo a 95% de confiança para o peso médio de uma cria ao desmame. Nestas circunstâncias, o estudo permite concluir que o peso médio de uma cria ao desmame é significativamente

- ☐ superior ou igual a 36, ao nível de confiança de 95%
☐ inferior ou igual a 42, ao nível de significância de 5%
☐ diferente de 42, ao nível de significância de 5%
☐ diferente de 35, ao nível de significância de 5%



+1/6/55+

Q18 Durante três meses submeteram-se 50 golden retriever a um tipo de dieta, tendo-se observado que a média das diferenças (inicial – final) do índice de massa corporal canino (IMCC) pré e pós-dieta é de 1.8 e o desvio padrão corrigido das diferenças é 0.3. Assuma que os dados são normalmente distribuídos. Averigüe, ao nível de significância de 1%, se o IMCC médio pré-dieta é significativamente superior ao IMCC médio pós-dieta.

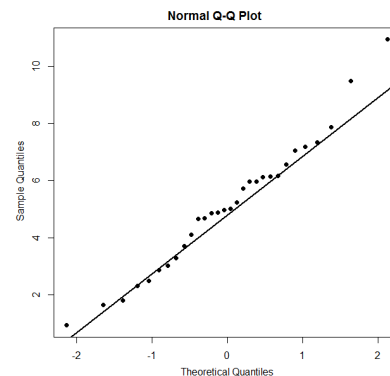
- ☐ Como valor p é inferior a 0.01, decide-se pela rejeição de H_0
- ☐ Como valor p é superior a 0.01, decide-se pela não rejeição de H_0
- ☐ Como valor p é inferior a 0.01, conclui-se que IMCC médio pré-dieta não é significativamente superior ao IMCC médio pós-dieta
- ☐ Como valor p é superior a 0.01, conclui-se que IMCC médio pré-dieta é significativamente superior ao IMCC médio pós-dieta

Q19 Foi recolhida uma amostra de valores correspondentes aos comprimentos dos corpos de 30 camarões de aquacultura selecionados casualmente. Assumindo que os dados são normais foi efetuado um teste- t para averiguar se o comprimento médio de corpos de camarões de aquacultura é significativamente inferior a 5.50. Sabe-se que o valor observado da estatística de teste T é $t_{obs} = -1.3$. Indique o valor p obtido. Tem-se

- ☐ valor $p = P(T \leq 5.50)/2 \approx 0.4999$
- ☐ valor $p = P(T \leq t_{obs}) \approx 0.10192$
- ☐ valor $p = P(T \leq 5.50) \approx 0.999997$
- ☐ valor $p = 2 * P(T \leq t_{obs}) \approx 0.20384$

Q20 Foi recolhida uma amostra de valores correspondentes aos comprimentos dos corpos de 30 camarões de aquacultura selecionados casualmente. Com esses valores foi construído o seguinte QQ-plot. Das seguintes afirmações, selecione a que não está correta.

- ☐ O QQ-plot da normal é uma ferramenta gráfica que permite averiguar se os dados provêm de uma população normal.
- ☐ Um QQ-plot da normal não é um teste de hipóteses associado a um dado nível de significância.
- ☐ É de crer que os dados provêm de uma população normal pois todos os pontos no QQ-plot se encontram relativamente próximos da reta.
- ☐ É de crer que os dados provêm de um modelo definido por uma reta.





+1/7/54+

Q21 De estudos ecológicos realizados em rios portugueses sabe-se que a concentração de estrôncio (mg/ml), no curso da água dos rios Sado e Guadiana, é bem modelado por distribuições normais. Foram recolhidas 50 amostras de cada rio. Das amostras do rio Sado registou-se uma concentração média de estrôncio de 38.1mg/ml com um desvio padrão corrigido de 1.2mg/ml. Da amostra do rio Guadiana registou-se uma média de 37 mg/ml com um desvio padrão corrigido de 1.1mg/ml. Averigue, ao nível de significância de 1% se a concentração média de estrôncio no rio Sado é superior à do rio Guadiana. Na sua resposta deverá: especificar as hipóteses H_0 e H_1 (0.7), averiguar a homogeneidade de variâncias (0.5), indicar o valor observado da estatística de teste (0.3), indicar o valor p do teste (0.4), decidir e concluir (0.6).

☐ 0 ☐ 0.7 ☐ 0.5 ☐ 0.3 ☐ 0.4 ☐ 0.6 ☐ 2.5



+1/8/53+

Mini Teste 4

Q22 Para averiguar o efeito da idade nos níveis de concentração sérica de Gentamicina no sangue, em ovelhas sujeitas àquele antibiótico, planeou-se a seguinte experiência envolvendo 18 ovelhas: 6 ovelhas do grupo etário 1 (idade entre 2 e 4 anos), 6 ovelhas do grupo etário 2 (idade entre 5 e 7 anos), e 6 ovelhas do grupo etário 3 (idade superior a 8 anos). A cada ovelha foi administrado o referido antibiótico na dose de 10 mg/kg de peso corporal e 2h depois foram recolhidos os níveis de concentração sérica (em g/ml) de Gentamicina no sangue. Com vista a efectuar uma ANOVA paramétrica foi construído o seguinte quadro resumo (incompleto):

Fonte de variação	Soma dos quadrados	Graus de liberdade	Media dos quadrados	Valor observado da estatística de teste	valor p
ENTRE GRUPOS	(a)	(c)	188.6	(g)	(h)
DENTRO DOS GRUPOS	265	(d)	(f)		
TOTAL	(b)	(e)			

Nas condições dadas tem-se que:

- ☐ (b) = 830.8 ; (c) = 3 ; (g) \approx 18.9.
- ☐ (b) = 359.3 ; (d) = 15 ; (g) \approx 10.7.
- ☐ (a) = 94.3 ; (f) = 17.7 ; (h) \approx 0.0013.
- ☐ (a) = 377.2 ; (d) = 15 ; (h) \approx 0.0013.

Q23 O planeamento considerado na questão **Q22** corresponde a uma

- ☐ ANOVA de efeitos fixos com 1 fator com 6 grupos.
- ☐ ANOVA de efeitos fixos a 1 fator e medidas repetidas.
- ☐ ANOVA de efeitos aleatórios com 1 fator.
- ☐ ANOVA de efeitos fixos com 1 fator com 3 grupos.

Q24 Para que a ANOVA paramétrica considerada na questão **Q22** seja válida, um dos pressupostos que deverá ser verificado é:

- ☐ a existência de homogeneidade de médias entre os grupos etários.
- ☐ que os dados dentro de cada grupo provenham de uma distribuição normal.
- ☐ a existência de homogeneidade de variância entre as ovelhas.
- ☐ a existência de igualdade de médias entre os grupos etários.



+1/9/52+

Q25 Foi aplicado o teste de Tukey de comparação múltipla aos dados considerados na questão **Q22**. Com recurso ao R foi obtido o seguinte resultado:

```
Tukey multiple comparisons of means
95% family-wise confidence level

            diff            lwr            upr            p adj
grupo etario 2-grupo etario 1 -8.666667 -14.969956 -2.363377 0.0073786
grupo etario 3-grupo etario 1 -10.500000 -16.803290 -4.196710 0.0016308
grupo etario 3-grupo etario 2  -1.833333  -8.136623  4.469956 0.7350194
```

Os resultados

- ☐ permitem concluir que, ao 5% de significância, os níveis médios de concentração sérica de Gentamicina no sangue não são significativamente diferentes entre os grupos etários 1 e 2.
- ☐ permitem concluir que, ao 5% de significância, os níveis médios de concentração sérica de Gentamicina no sangue são significativamente diferentes entre os grupos etários 2 e 3.
- ☐ não são válidos pois não é possível aplicar o teste de Tukey a este conjunto de dados.
- ☐ permitem concluir que, ao 5% de significância, os níveis médios de concentração sérica de Gentamicina no sangue são significativamente diferentes entre os grupos etários 1 e 3.

Q26 São conhecidos 3 tipos diferentes de habitat de porco preto ibérico (*Sus scrofa mediterraneus*). Pretende-se investigar se o habitat afeta o peso médio (μ) dos porcos pretos ibéricos de uma determinada faixa etária. Para tal foram recolhidos, em cada habitat, os pesos de 5 porcos pretos ibéricos.

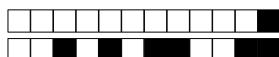
$h1$	$h2$	$h3$
45	72	65
53	75	61
57	85	51
48	81	55
60	74	63

Assuma válidos os pressupostos da ANOVA paramétrica. Neste planeamento, avaliar a questão de investigação corresponde a testar:

- ☐ $H_0 : \mu_{h1} = \mu_{h2} = \mu_{h3} = 0$ vs $H_1 : \mu_{hi} \neq 0$ para algum habitat hi .
- ☐ $H_0 : \mu_{h1} = \mu_{h2} = \mu_{h3} = \mu$ vs $H_1 : \mu_{hi} \neq 0$ para algum habitat hi .
- ☐ $H_0 : \mu_{h1} = \mu_{h2} = \mu_{h3} = \mu$ vs $H_1 : \mu_{hi} \neq \mu_{hj}$ para algum par de habitats hi, hj ($i \neq j$).
- ☐ $H_0 : \mu_{h1} = \mu_{h2} = \mu_{h3} = 0$ vs $H_1 : \mu_{hi} \neq \mu_{hj}$ para algum par de habitats hi, hj ($i \neq j$).

Q27 Realizando o teste estatístico referido na alínea anterior, com base no conjunto de dados indicado, tem-se que o valor observado da estatística de teste é

- ☐ ≈ 5.8195
- ☐ ≈ 5.827
- ☐ ≈ 0.00006
- ☐ ≈ 24.47



+1/10/51+

Q28 Um investigador pretende averiguar se a concentração média de cálcio no sangue depende da realização um dado tratamento hormonal e do sexo. Para tal, foram selecionados casualmente 20 machos e 20 fêmeas de uma população de mamíferos. Dos 40 mamíferos foram selecionados casualmente 20 (10 machos e 10 fêmeas) aos quais foi administrado o tratamento hormonal. Aos restantes (10 machos e 10 fêmeas) foi administrado um tratamento placebo. Ao fim de um mês foi registado a concentração de cálcio dos 40 mamíferos em observação. Assuma a validade dos pressupostos da ANOVA paramétrica. Com recurso ao R obteve-se o seguinte resultado:

```
          Df Sum Sq Mean Sq F value Pr(>F)
tratamento  1  0.1796  0.17956    7.110  0.0114
sexo         1  0.1323  0.13225    5.237  0.0281
tratamento:sexo  1  0.0314  0.03136    1.242  0.2725
Residuals   36  0.9092  0.02526
```

Assim, podemos concluir que, ao nível de significância de 5%, o tratamento hormonal afeta significativamente a concentração média de cálcio no sangue de um mamífero já que o valor p associado ao teste é

- ☐ 0.0314
- ☐ 0.2725
- ☐ 0.1796
- ☐ 0.0114

Q29 Relativamente à alínea anterior o valor 0.2725 que surge nos resultados obtidos do R significa que, ao nível de significância de 5%,

- ☐ existe uma concentração média de cálcio significativamente diferente entre o sexo e o tratamento hormonal.
- ☐ não existe interação significativa entre o tratamento hormonal e o sexo na concentração média de cálcio no sangue.
- ☐ não existe evidência estatística do sexo afetar significativamente o tratamento hormonal.
- ☐ não se rejeita a hipótese nula de igualdade de médias.

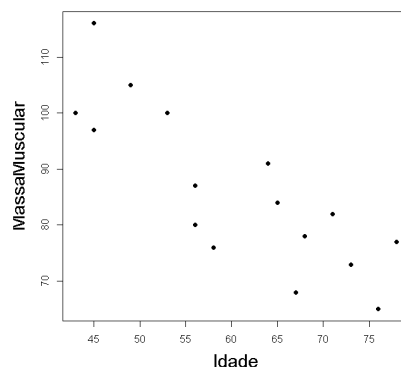


+1/11/50+

Mini Teste 5

Q30 Para averiguar a existência de relação entre a massa muscular de um adulto com a sua idade, um nutricionista recolheu uma amostra de 18 indivíduos, com idade entre 40 e 79 anos, e observou em cada um deles a idade (X) e a massa muscular (Y). Com os dados recolhidos obteve o seguinte diagrama de dispersão. Logo, é possível concluir que

- ☐ um adulto mais velho tenderá a ter menor massa muscular.
- ☐ existe uma associação do tipo linear entre os 18 indivíduos e cada uma das variáveis (X e Y).
- ☐ o valor do coeficiente de correlação amostral de Spearman entre X e Y deve ser muito próximo de zero.
- ☐ o coeficiente de correlação amostral de Pearson entre X e Y deve corresponder a um valor negativo indicando que os adultos mais novos tendem a ter menor massa muscular.



Q31 Tomando a amostra considerada na questão **Q30**, e usando o coeficiente de correlação de Pearson, foi testado se as duas variáveis X (idade) e Y (massa muscular) estão correlacionadas. Foi obtido o valor p do teste igual a 1.5×10^{-5} . Podemos então concluir que

- ☐ a idade e a massa muscular estão relacionados significativamente por uma função linear.
- ☐ a massa muscular e a idade estão relacionadas por uma constante igual a 1.5×10^{-5} .
- ☐ o coeficiente de correlação de Pearson entre a idade e a massa muscular não é significativamente diferente de zero.
- ☐ existe uma relação de independência estatisticamente significativa entre a idade e a massa muscular.

Q32

O nutricionista propôs-se realizar uma análise de resíduos para verificar os pressupostos de realização de uma análise de regressão linear simples. Assim, usando os resíduos (erros) deverá verificar que

- ☐ o declive da reta de regressão a ajustar aos dados é nulo.
- ☐ os dados não provêm de uma população normalmente distribuída.
- ☐ os erros têm variância constante.
- ☐ os erros têm média positiva.

Q33 Um investigador pretende avaliar se uma dada balança está bem calibrada. Considerou 9 pesagens conhecidas (X) e determinou o respetivo peso atribuído pela balança (Y). Os dados foram:

X	2.0	2.0	2.0	4.0	4.0	4.0	6.0	6.0	6.0
Y	2.1	1.8	1.9	4.5	4.2	4.0	6.2	6.0	6.5

A equação da reta de regressão estimada com base nos dados é

- ☐ $y = 0.202 + 0.919x$
- ☐ $y = -0.167 + 1.075x$
- ☐ $y = 0.919 + 0.202x$
- ☐ $y = 1.075 - 0.167x$



+1/12/49+

Q34 Tomando os dados considerados na questão **Q33**, o investigador verificou que a relação entre o peso real e o peso atribuído pela balança é bem modelada por uma regressão linear simples já que o valor obtido para o coeficiente de determinação é elevado e aproximadamente igual a:

- ☐ 1.075.
☐ 0.988.
☐ 0.994.
☐ 23.72.

Q35 Tomando os dados considerados na questão **Q33**, foi realizado o teste ao declive do modelo de regressão para avaliar se é positivo, ao nível de significância de 5%. Tal corresponde a ter:

- (A) $H_0 : \beta_1 = 0$ vs $H_1 : \beta_1 > 0$
 (B) Valor observado da estatística de teste aproximadamente igual a 0.222
 (C) Valor p do teste aproximadamente igual a 0.00000003

Podemos dizer que as afirmações

- ☐ (B) e (C) estão certas mas (A) está errada. O declive é significativamente positivo.
☐ (A) e (B) estão certas mas (C) está errada. O declive não é significativamente positivo.
☐ (A), (B) e (C) estão certas. O declive não é significativamente positivo.
☐ (A) e (C) estão certas mas (B) está errada. O declive é significativamente positivo.

Q36 Tomando os dados considerados na questão **Q33**, qual a alteração esperada no peso indicado pela balança quando o verdadeiro peso sofre um aumento 1.5 unidades?

- ☐ sofre uma diminuição de aproximadamente 0.167 unidades
☐ sofre um aumento de aproximadamente 1.446 unidades
☐ sofre um aumento de aproximadamente 1.075 unidades
☐ sofre um aumento de aproximadamente 1.613 unidades

Q37 Do seguinte quadro pretende-se avaliar se a escolha de um sabor de gelado depende do género.

	Chocolate	Baunilha	Morango
Masculino	100	120	60
Feminino	350	200	90

É verdade que:

- ☐ Se as variáveis forem independentes, é esperado observar 100 pessoas do sexo masculino e que preferem o sabor de chocolate.
☐ A escolha do sabor de gelado não depende do sexo, considerando 5% de significância.
☐ O valor observado da estatística de teste é aproximadamente 28.4.
☐ Existem 4 graus de liberdade no teste.